

Essays on competitive attitude

Citation for published version (APA):

Da Cunha Batista Geraldles, D. (2018). *Essays on competitive attitude: sex, stereotypes, and group membership*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20180509dg>

Document status and date:

Published: 01/01/2018

DOI:

[10.26481/dis.20180509dg](https://doi.org/10.26481/dis.20180509dg)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Essays on Competitive Attitude

Sex, stereotypes, and group membership

© Diogo Geraldes, Maastricht 2017
<https://sites.google.com/site/diogogeraldes>

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without prior written permission from the author.

Requests for permission to reproduce materials from this publication should be sent via email to diogo.geraldes@gmail.com.

Published by Maastricht University
Printed by Datawyse – Universitaire Pers Maastricht
Cover design by Sophia de Groot
ISBN: 978 9462 95 902 6

Essays on Competitive Attitude

Sex, stereotypes, and group membership

DISSERTATION

to obtain the degree of Doctor at

Maastricht University

on the authority of the Rector Magnificus, Prof. dr. Rianne.M. Letschert

in accordance with the decision of the Board of Deans,

to be defended in public

on Wednesday, 9th of May 2018, at 14.00 hours

by

Diogo Geraldés

Supervisors:

Prof. Dr. Arno Riedl

Dr. Martin Strobel

Assessment Committee:

Prof. Dr. Alexander Brüggen

Dr. Thomas Buser

Prof. Dr. Nagore Iriberrí

Dr. Christina Rott

To my father

Contents

Chapter 1. Introduction to the thesis

1.1. Relevance of behavioral and experimental economics	1
1.2. Objectives of the thesis	6
1.3. Overview of Part I: Gender competition	6
1.4. Overview of Part II: Group membership	8
References	10

PART I | GENDER COMPETITION

Chapter 2. Women dislike competing against men

2.1. Introduction	13
2.2. Design of the experiment	19
2.3. Basic experimental results	23
2.4. Stereotypes, confidence, and the “irrational” competitors	25
2.5. Uncovering the beliefs about the sex of potential competitors	33
2.6. Discussion of policy implications	38
2.7. Concluding remarks	40
References	43
Appendices	45

Chapter 3. Sex and performance under competition: Is there a stereotype threat shadow? An intriguing behavior of men

3.1. Introduction	67
3.2. Design of the lab experiment	70
3.3. Hypotheses	75
3.4. Results of the lab experiment	78
3.5. Effort provision—A normative view	92
3.6. Discussion	95
3.7. Concluding remarks	96
References	99
Appendices	102

PART II | GROUP MEMBERSHIP

Chapter 4. Group membership and individual competitive behavior

4.1. Introduction	135
4.2. The experiment	140
4.3. Competition level results	144
4.4. Real-effort task performance results	153
4.5. Concluding remarks	156
References	158
Appendices	160

Chapter 5. Synopsis of the thesis

5.1. Women dislike competing against men	189
5.2. Sex and performance under competition: Is there a stereotype shadow? An intriguing behavior of men	190
5.3. Group membership and individual competitive behavior	190

Valorization Addendum	193
------------------------------	-----

Academic Biography	195
---------------------------	-----

Chapter 1

Introduction to the Thesis

“Doing experimental economics has changed the way I think about economics. There are many reasons for this, but one of the most prominent is that designing and conducting experiments forces you to think through the process rules and procedures of an institution. Few, like Einstein, can perform detailed and imaginative mental experiments. Most of us need the challenge of real experiments to discipline our thinking.”

Vernon L. Smith’s Nobel Prize Lecture (2002)

1.1. Relevance of behavioral and experimental economics

Broadly speaking, the research output of this thesis lies in the fields of behavioral and experimental economics. Thus, particularly directed to the unfamiliar reader, I start by briefly putting into perspective the scientific relevance of these fields. The other readers may perhaps prefer to start their reading in section 1.2..

1.1.1. A novel methodology

A general Google search that asks, “What is behavioral economics”, yields a dictionary definition describing behavioral economics as “a method of economic analysis that applies psychological insights into human behaviour to explain economic decision-making”. In my view, however, the popular conception of behavioral economics (BE) as economic analysis with insights from psychology understates the state-of-the-art of research in BE. Psychology has certainly been

the most influential (outside economics) discipline of BE thus far. But, in the recent years, other scientific fields such as anthropology, cognitive neuroscience, evolutionary biology, and sociology are conspicuously influencing BE as well. In fact, its multiplicity of roots and influences have probably turned BE into one of the most vibrant interdisciplinary fields.

Succinctly, BE is a sub-field of economics that takes into consideration findings, models, and methods from several other scientific disciplines to investigate human individual and social behavior in economically relevant settings. Thus, as a natural consequence of its approach, BE postulates a variety of alternatives to long-established paradigms in economics such as, for example, the rational choice model of behavior.

Predominantly combined with BE is experimental economics (EE), which has also become a separate subfield of economics. EE is a methodology of empirical investigation. Essentially, it consists of collecting data in a controlled environment using a toolbox with distinctively valuable features for economic analysis.

For the sake of brevity, I do not compile an exhaustive list of the advantages linked to EE methods, but rather refer only to the ones that I deem the most important. For the sake of relevance, I focus on methods used in laboratory experiments because this is the type of experiment underlying the research output that I report in this thesis.¹

Relatively to other empirical methods used in economics, the most important feature of economic lab experiments is **enhanced control**. Among others, this feature permits economists:

- To investigate *causal* effects because the experimenter can perfectly set the exogenous and endogenous variables of interest ex-ante. This aspect rules out the problem of *reverse causality*, which often undermines the interpretation of empirical results.
- To perform *ceteris paribus* changes in an exogenous variable. This aspect makes economic experiments a very convenient setting to examine comparative statics empirically.
- To randomly assign participants to treatment conditions. This aspect rules out the problem of *selection bias*, which is usually present when using other empirical methods.

Relative to experimental practices in other social sciences, two noteworthy features of economic experiments are the use of **monetary incentives** and the

¹ *Field experiments* are also commonly conducted in EE. Although many of the advantages of lab experiments are also present in field experiments, there are key differences. On the one hand, lab experiments offer more control. On the other hand, field experiments have a higher external validity. To make the most of the two methods, there is an increasing wave of advocates for the parallel running of lab and field experiments as well as *lab-in-the-field* experiments.

No-deception rule. These features particularly contrast with experiments in psychology.² The advantages of these two features are as follows:

- Monetary incentives can be aligned with the variable of interest so that participants can attain the best monetary outcome only if they behave according to their true preferences. Thus, using *incentive compatible* monetary mechanisms means that participants have no incentive to lie when asked about their preferences and beliefs, and have no incentive to shirk when asked to perform a real-effort task.
- Instituting the no-deception rule allows experimenters to make credible the truthfulness of the information they provide in experiments. This procedure should lead participants to believe in the experimental instructions rather than trying to outguess the experimenter. In short, the desired effect of using the no-deception rule should emerge because we are essentially facing an issue of *reputation in a repeated game*.³

A last noteworthy feature of economic lab experiments is their ***de facto* replicability**. In principle, any lab experiment is replicable. In practice, however, this is unrealistic for much of the experimental work in the social sciences because protocols, instructions, computer codes, and data are not publicly or readily available. In EE, this hurdle is tackled via the prospect of publishing as a discipline device. Specifically, publishing experimental work in top economics journals requires the online posting of the experimental procedures and data. This provides an incentive for experimenters to strictly follow reported procedures. Perhaps not surprisingly, comparing the recent replication study regarding EE work (Camerer et al. 2016) to the recent Reproducibility Project Psychology (Open Science Collaboration 2015), we conclude that lab experimentation in economics generates better replicability of results.

1.1.2. Purposes

The combination of behavioral and experimental economics allows economists to carry out multiple scientific purposes. Without intending to be exhaustive, studying human behavior via economic lab experiments:

- Provides a unique framework to **test economic theories**.

² In fact, the American Psychological Association Code of Ethics states that psychologists do not conduct studies involving deception unless effective nondeceptive alternative procedures are not feasible. In my view, the recurrent observation of deception in psychology studies is thus an issue of lack of enforcement since it is usually easier to design experiments using deception.

³ During my (early stage) career in the field of EE, I already had the privilege of meeting innumerable experimental economists and visit several experimental economics labs. Even though I never met a single experimental economist that would say not to follow the no-deception rule, I was surprised by the fact that many do not make the rule common knowledge to their subject-pool nor refer in the experimenter's welcome speech that participants can inspect the veracity of specific information at the end of the experiment. I deem these two elements fundamental to attain the scientific advantage of using the no-deception rule.

- Allows the **search for empirical regularities**, which offers an excellent basis for new theories.
- Provides a clean method for the **elicitation of preferences**.
- Allows **“wind tunnel” experimentation**, which is extremely useful to advise policymaking.
- Can play a significant role in the **education of economics**.

I particularly want to highlight the last point because the educational purpose of economic experiments has been highly neglected hitherto. As an enthusiastic teacher and experimentalist in economics, I have been making use of this tool in my classes, and the results have been extraordinary. Economic concepts can be too abstract to provide students a grasp of how to apply them. Conducting small-scale classroom economic experiments has considerably improved my students' understanding of key concepts. I think classroom experiments facilitate students' learning process because they appeal to their visual memory. In short, in addition to their value in research, I strongly advocate the use of experimental methods in the teaching of economics.

1.1.3. Research avenues and applications

Some prominent avenues of research in which behavioral and experimental economics have advanced the economics science include the understanding of: i. **Decision under uncertainty** (e.g., risk preferences); ii. **Intertemporal choice** (e.g., hyperbolic discounting model); iii. **Social preferences** (aka other-regarding preferences) such as altruism, reciprocity, inequality aversion, trust, spite, and guilt aversion; iv. **Behavioral game theory** (e.g., level-k reasoning) and **networks**, and v. **Learning** (e.g., reinforcement learning).

The applications of the knowledge generated by behavioral and experimental economics are far-reaching. This knowledge applies to domains such as, for example, consumer behavior, labor market supply, self-control problems (e.g., unhealthy eating), stock market trading, pension enrollment, organ donation, provision of public goods, and energy supply.

1.1.4. Critiques

As recurrently observable in the history of science, new approaches incur skepticism. Behavioral and experimental economics are no exception.

Some notorious objections to the use of economic experiments are that they **lack realism**. That is, critics posit that such experiments are artificial; and **lack representativeness** because participants are students. My stand on these two critiques is that they are fallacious.

Congruent with the argument used in Charles Plott (1982), economic experiments are conducted with real people who make real decisions and make real money under real rules. Moreover, we can perfectly add realism in controlled

steps to simplified experimental settings. But in fact, experimental economists purposely design simplified experimental settings. As with theoretical economic models, the more parsimonious an economic experiment is, the better. The kernel of economic experiments rests on designing simple special cases relevant to the understanding of the complex, and not the other way around. In short, economic experiments are not artificial; it is precisely because they are real that economic experiments are informative.

Also, economic experiments are not conducted only with students, as many critics emphasize. In fact, many experiments in economics are conducted with non-student participants (see Kagel and Roth 2016).⁴ Moreover, in general, the conclusions reached in experiments with students compared to experiments with other groups are similar (Fréchette 2015).

A prominent objection to EE work that I too acknowledge relates to the **induction problem** (Hume 1748). This problem can undermine the **external validity** of economic lab experiments. Nevertheless, in my view, this is not an argument against conducting economics experiments altogether. To illustrate, let us consider the classical *sunrise problem* as a counterargument to this possible shortcoming.

For billions of years, the sun has been rising every morning. According to the induction problem, we cannot make the inference that tomorrow morning the sun will rise again. In my view, however, it is fundamental to distinguish between strong and weak inductive statements. If many experiments show that under a similar set of conditions the same empirical regularities emerge, the likelihood that the same empirical regularities will occur under similar circumstances is higher than if we based our prediction in only one experiment. Returning to the classical *sunrise problem*, nowadays we even know with certainty that the sun will eventually burn out, which means that we already know that the conclusion based on induction that the sunrise will take place every morning is false. Nevertheless, before the sun burns out, does the induction problem nullify the insight gained through induction regarding the predictability of the sun rising tomorrow, next week, or next year?

A second argument against the objection of the induction problem derives from Karl Popper's work (Popper 1959). According to Popper, we cannot prove a theory through empirical evidence, regardless of the number of observations we obtain. Thus, the induction method is flawed in its departure point—it poses the wrong goal of *proving* a theory. Conversely, we only need one disconfirming

⁴ Rather than the unrepresentativeness of students, a significant discussion in the behavioral sciences is whether results obtained with samples (either consisting of students or other groups) drawn from Western, Educated, Industrialized, Rich, and Democratic societies are representative of the *Homo sapiens*' behavior in general. For instance, Henrich, Heine, and Norenzayan (2010) find that this subpopulation of humankind—in which most of the behavioral claims published in top journals are routinely based on—is highly unrepresentative of the species. Perhaps not innocuously, the authors use the acronym WEIRD to dub people from these societies.

observation to conclude that a theory is not true. Thus, Popper's argument implies that empirical work in science is fundamentally necessary for attempting to falsify existing theories.

All in all, hostility towards the use of economic experiments has faded with time. While experimental work in economics has been regarded with much skepticism for a considerable amount of time (particularly among *mainstream* economists), running experiments is currently widely regarded as part of economists' toolbox to explain and describe economic phenomena.

In sum, based on a genuinely interdisciplinary approach, behavioral and experimental economics are expanding economics science in its unique fashion by generating alternative theoretical insights, improving the testing of economic predictions, and providing invaluable assistance to tailor better policy.

1.2. Objectives of the thesis

In this thesis, I pursue two fundamental objectives. Firstly, through the combination of economic experiments and economic theory, I aim to understand motivations and incentives underlying behavior in economically relevant **competitive** settings. Secondly, I aim to use the behavioral insights from my research to derive policy implications because I am firmly convinced that my type of research is valuable for the design of welfare-enhancing policies.

My work can be applied, for instance, to design policy based on *nudge theory*, which advocates we can induce better decisions without constraining freedom of choice if we properly design the choice architecture (see Thaler and Sunstein 2008). Policy makers are increasingly open to the idea of *nudging* people, as is evident from the recent OECD and World Bank endorsements for the use of behavioral insights to design policy (OECD 2017; World Bank 2015).

The thesis consists of two parts. In the first part, I investigate gender competition and its policy implication for the labor market. In the second part, I investigate the impact of group membership on individual willingness to compete. In the following two sections, I overview both of these endeavors.

1.3. Overview of Part I: Gender competition

A politically, socially, and economically important stylized fact about gender differences is the observable gap in positions in the workplace. In particular, the underrepresentation of women in high-level decision-making positions has received much attention in political and business debate and among researchers from diverse fields.

Several *raison d'être* have been proposed to explain the gender gap in positions in the workplace. Examples have pertained to discrimination against women and women's higher sensitivity to work-family conflict. More recently, behavioral

economists advanced an alternative explanation: women have a weaker preference to compete compared to men (see Niederle and Vesterlund 2007).

The first part of my dissertation, which consists of two chapters, focuses on the study of men's and women's competitive attitudes from a new perspective, and if (and how) this perspective could shed light on the underrepresentation of women in male-typed labor markets.

1.3.1. Overview of Chapter 2

In Chapter 2, I explore the relevance of the sex of a competitor in explaining men's and women's willingness to compete.

To this end, I conduct an innovative laboratory experiment to examine whether beliefs about the sex of potential competitors affects men's and women's decision to enter competition. Specifically, I test whether women have a weaker preference to compete per se, or whether they rather shy away from competing against men.

The results support the latter hypothesis. Moreover, only women are sensitive to the different cues I provide concerning the sex of potential competitors, and their competitiveness is largely driven by their beliefs about other women's competitive attitude.

These findings allow me to distill distinctive policy implications for the labor markets in which women are underrepresented. Namely, I advocate the removal of information that highlights women underrepresentation in advertisements of a male-typed job. Instead, I propose the inclusion of persuasive references to female role models in the job.

1.3.2. Overview of Chapter 3

In Chapter 3, I focus on a different strand of the gender competition literature.⁵ Rather than self-selection into a competition, I study men's and women's competitive performances under an exogenously set mixed-sex competition. Specifically, I conduct laboratory and online experiments to investigate the impact of stereotype threats on men's and women's performances.

In this chapter, I also study the effort men and women should strategically exert in the competition, which consists of a pairwise winner-take-all tournament, but with an endogenous prize.

I find no significant gender difference in competitive performance when I induce stereotype threat against women. However, men's competitive performance is negatively affected when I induce stereotype threat against men.

⁵ This chapter is based on Geraldes, Riedl, and Strobel (2016)

My experimental and theoretical results have practical implications regarding policy design to cope with stereotype threat at the workplace. In stark contrast to the psychology literature on stereotype threat, which advocates “stereotype nullification” (i.e., to explicitly provide individuals with information that does not conform to the stereotype) to prevent the adverse effect of stereotype threat in performance, my results indicate that no information should be provided at all. That is, if men and women are already competing against each other, my findings suggest they cope well with a gender stereotype they hold concerning performance in a mixed-sex competition.

1.4. Overview of Part II: Group membership

In the second part of my dissertation, I investigate the relationship between group membership and individual competitive attitude. Although competitive settings have been studied from many different perspectives, a noteworthy element that research to date has been neglecting is the effect of group membership on competitive attitude. Promotions in labor environments, political processes, or sport competitions are examples of the many situations that can be modeled as a tournament in which group membership manifestly plays a relevant role.

To tackle this challenge, I draw inspiration from the blossoming economics literature on *social identity*. Based on a central paradigm in social psychology—*Social Identity Theory* (Tajfel and Turner 1979)—, this recent strand of literature has elegantly challenged the standard assumption in economic theory that individual behavior is determined solely at the individual level (for a theoretical challenge, see Akerlof and Kranton 2000; for an experimental challenge, see Charness, Rigotti, and Rustichini 2007).

1.4.1. Overview of Chapter 4

In Chapter 4, I test if (and how) group membership affects individual competitive behavior. To this end, I conduct a laboratory experiment in which I elicit participants’ competitive preferences in two distinct conditions. In one condition, the participants are induced with group membership whereas. In a second condition, participants act on their own. Moreover, to elicit participants’ competitive preferences, I devise a new indicator that allows for the measurement of several different levels of competitiveness.

I find that group membership restrains competitive behavior. In short, a participant’s choice of competition level in my setting is an individual decision. However, for the participants induced with group membership, the choice of competition level affects the payoff of the other members of their group. Supported by participants’ second-order beliefs about the choice of competition level, I argue that group membership restrains competitive behavior in my setting (partially) because of guilt aversion.

Moreover, even though members of a group choose lower competitive levels, I find that they perform better under competition relative to control participants. I attribute the latter result to the presence of an audience during the competition, which is composed of the other members of the group in the group membership context.

Naturally, more research is needed to evaluate the robustness of these results. Nevertheless, these findings suggest two noteworthy policy implications: i. To deter reckless decisions in competitive labor market: Index the bonus of workers to the performance of their peers; ii. To improve workers' productivity, we should strengthen their group spirit to compete against external organizations rather than promoting internal competition.

REFERENCES

- Akerlof, G., & Kranton, R. (2000). Economics and identity. *Quarterly Journal of Economics*, 115(3), 715-53.
- Camerer, C., Dreber, A., Forsell, E., Ho, T., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433-36.
- Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual Behavior and Group Membership. *American Economic Review*, 97(4), 1340-52.
- Fréchette, G. (2015). Laboratory Experiments: Professionals versus Students, in *Handbook of Experimental Economic Methodology*, Guillaume R. Fréchette and Andrew Schotter (editors), Oxford University Press, pp: 360-390.
- Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 62–135.
- Hume, D. (1748). *An enquiry concerning human understanding*. (original version); Hume, D., & Millican, P. (2007). *An enquiry concerning human understanding*. Oxford: Oxford University Press. (consulted version)
- Kagel, J., & Roth, A. (2016). *The Handbook of Experimental Economics, Volume 2*. ed. Princeton University Press.
- Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics*, 122(3), 1067-1101.
- OECD (2017). *Behavioural Insights and Public Policy report*.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Plott, C. (1982). Industrial organization theory and experimental economics. *Journal of Economic Literature*, 20(4), 1485-1527.
- Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.
- Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In *The social psychology of intergroup relations* (pp. 33-47). ed. S. Worchel; Austin, W.. Monterrey CA: Brooks/Cole.
- Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Connecticut (USA): Yale University Press.
- World Bank (2015). *Mind, Society, and Behavior report*.

Part I

Gender Competition

Chapter 2

Women dislike competing against men¹

2.1. Introduction

“Let me tell you about my trouble with girls...Three things happen when they are in the lab: You fall in love with them, they fall in love with you, and when you criticize them they cry”.² These recent controversial remarks of the Nobel Laureate Tim Hunt anecdotally illustrate a pervasive phenomenon in the workplace; men often employ gender stereotypes to look-down-upon women. Concern over gender stereotypes is growing, as attested by the massive attention paid to recent sexism cases involving major tech firms such as Facebook and Twitter. While the demand side of the market has already started to address gender problems at the workplace related to stereotypes (e.g., McKinsey’s gender diversity program), and some research has explored its effects from a demand side perspective (e.g., Reuben, Sapienza, and Zingales 2014), how gender stereotypes affect the supply side is unknown. In this chapter, we study whether (and how) women’s beliefs about the sex of potential competitors influence their decision to enter a competitive male-typed domain.³

¹ This chapter is written using the *pluralis majestatis* (also known as the royal “we”) and makes use of the *Singular they* as a gender-neutral singular pronoun.

² Remarks of Tim Hunt at the 2015 World Conference of Science Journalists.

³ Throughout this chapter, male-typed refers to domains/situations/contexts/environments/areas/fields in which men are stereotypically believed to be more skilled and/or overrepresented.

The seminal work of Niederle and Vesterlund (2007) has catalyzed a host of studies investigating men's and women's self-selection into competitive environments (for a review of the literature, see Niederle and Vesterlund 2011). The prevalent procedure to measure competitive preference across these studies is to ask participants to choose between a non-competitive and a competitive payment scheme for a subsequent performance of a real-effort task. The main finding is that men are significantly more inclined to choose the competitive payment. Accordingly, it has been suggested that a possible reason for the well-documented gender gap in wages and positions in the workplace as well as in educational choices is that men have a stronger preference to compete compared to women.⁴ An open key question, however, is whether the accumulating evidence of women's lower inclination to compete is due to an aversion to competition per se, or because women dislike competing against men. The objective of this chapter is to answer this question by examining women's beliefs about the sex of potential competitors, a presumably important element influencing their entry into competition.

A common procedure in many studies that report women's lower inclination to choose the competitive option is to conduct mixed-sex laboratory sessions without referencing the gender composition of participants (e.g., Niederle and Vesterlund 2007; Gneezy, Leonard, and List 2009; Cason, Masters, and Sheremeta 2010; Dohmen and Falk 2011; Wozniak, Harbaugh, and Mayr 2014; Almås, Cappelen, Salvanes, and Tungodden 2016; Buser, Dreber, and Mollerstrom 2016). Hence, if female participants perceive the gender composition in the lab and, accordingly, correctly anticipate potential male competitors, then the claim that women dislike competition per se might be compromised. If women dislike competing against men, the stylized fact that women shy away from competition might be a result of the mixed-sex composition in the lab rather than an intrinsic weaker preference for competition. Thus, for drawing accurate policy implications, elucidating the relationship between women's willingness to compete and their beliefs about the sex of potential competitors is of utmost importance.

Our study has two primary purposes. First, we conduct a controlled laboratory experiment to investigate whether women have a weaker preference to compete per se, or rather just shy away from competition against men. To this end, we measure men's and women's self-selection into competition across conditions entailing different cues for the gender composition of potential competitors. Specifically, we ask each participant to choose between a non-competitive and a competitive payment scheme for a subsequent performance of a mathematical task in three conditions: i. *Replication condition*: the gender composition of participants present in the lab is symmetric; ii. *Choice of sex condition*: the gender

⁴ For instance, although the percentage of women CEOs in the Fortune 500 companies is at an all-time high of 6.4% (June 2017), the figure is still dismal for women (source: fortune.com). For further evidence on gender unequal outcomes at the workplace see, e.g., Blau, Ferber, and Winkler (2013). For evidence on gender differences in educational choice and its correlation with competitiveness see Buser, Niederle, and Oosterbeek (2014).

composition of participants present in the lab is symmetric, but a participant can choose the sex of the competitor while deciding to compete; iii. *All women condition*: only female participants are present in the lab. Our conjecture is that women's beliefs about the sex of potential competitors influence their decision to enter competition. In particular, we hypothesize that women are not less competitive than men but, at least in male-typed domains, they dislike facing a male competitor. The second purpose of this chapter is to assess if (and how) women's willingness to compete in the three conditions could help us to tailor policy intervention to encourage women to compete.

A central element of our experiment is that we do not inform participants about the gender composition in the lab before eliciting their choice of payment scheme, but rather rely on participants' perception of the gender composition, which we elicit at a later moment in the experiment. With this procedure, we aim not only to minimize the chance of participants realizing the goal of the experiment but also to better capture real-world decisions as, for instance, deciding whether to apply for a job. Moreover, the focus of this chapter is on a male-typed domain because the most representative labor markets and educational programs in which the gender gap is a serious concern (i.e., high-level business positions, STEM fields) are stereotypically believed to be male-typed domains.⁵ To create a male-typed environment in the lab, we choose a mathematical task because we expect female participants to be aware of the well-known stereotype that men are better at mathematics, and there is an empirical and theoretical basis to expect women to dislike competing against men in this situation.

Psychology research shows evidence that stereotype threat triggers women to feel apprehensive of the possibility of being judged based on negative stereotypes about their group in a (non-competitive) mixed-sex math context (e.g., Spencer, Steele, and Quinn 1999). Stereotype threat is defined as a situational predicament where an individual is at risk of confirming a negative stereotype about their group. Performance impairment in situations where an individual's poor performance conforms to the stereotype is the effect mostly associated with stereotype threat (for a review of the literature, see Inzlicht and Schmader 2013; Spencer, Logel, and Davies 2016). Therefore, stereotype threat theory might explain the influential results of Gneezy, Niederle, and Rustichini (2003). While investigating exogenously generated competition within a male-typed domain, these authors show that men and women perform equally well in single-sex groups, but women perform worse when competing in mixed-sex groups.

A conceivable side effect of past repetitive exposure to math stereotype threats is for women to anticipate a poor math performance when on the verge of a

⁵ We could argue that it is interesting per se to investigate whether the stylized result that women shy away from competition also holds in a female-typed domain. In fact, because the seminal work of Niederle and Vesterlund (2007) as well as many of the follow-up studies that replicate their finding use a male-typed work task, some studies have already shown that the gender gap in competition entry is reduced when a female-typed work task is used instead (e.g., Kamas and Preston 2012).

mixed-sex competition. Accordingly, a reason for expecting women to shy away from competition against men in a math context could simply be a decrease in their confidence level. Bordalo, Gennaioli, and Shleifer's (2014) self-stereotyping model formally describes a confidence mechanism that predicts women's lower inclination to enter a math competition. Specifically, their model assumes that when a woman compares herself to men in a situation involving a math competition, she underestimates her probability to win because the stereotype that men are better at math amplifies believed differences in ability between the sexes. Therefore, the model predicts that women will enter competition less often when their confidence level is biased by negative self-stereotyping.

We consider a complementary reason for expecting women to shy away from competition against men in a math context. If women are prone to a math stereotype threat and they are chronically exposed to this threat, a logical reaction of women may be to develop a preference to avoid situations where their math ability can be compared to men's. Accordingly, women should be particularly susceptible to dislike situations involving a math competition against men because the mixed-sex competition should heighten women's awareness of gender, thus amplifying the situational predicament of stereotype threat.

The closest previous work to our study is scant. Gneezy et al. (2009) acknowledge the possible importance of female participants deducing the gender distribution of potential competitors⁶, and Niederle, Segal, and Vesterlund (2013) recognize that quotas make the competition more sex-specific. To our knowledge, however, only two studies have attempted to manipulate the gender composition while investigating entry into competition. Using mixed- and single-sex groups, Booth and Nolen (2012) study boys and girls from either coeducational or single-sex schools. Their results suggest that girls are more willing to compete against other girls, but the figure is not segmented by the type of education. Gupta, Poulsen, and Villeval (2013) conduct balanced mixed-sex sessions in which they randomly match participants in mixed- and single-sex pairs. The authors find that women's choice of the competitive payment does not depend on the sex of their pair. However, in their setting, a participant is automatically the winner in case a participant is the only one to choose competition. Thus, the choice of payment scheme also depends on outguessing whether a participant's pair chickens out. Consequently, we cannot disentangle whether the result is mainly driven by women's competitive preference or outguessing. Hence, despite these two attempts⁷, the connection between a woman's decision to enter competition and

⁶ Using an elegant experiment, Gneezy et al. (2009) find that women of a matrilineal society in India (Khasi tribe) and men of a patriarchal society in Tanzania (Maasai tribe) enter competition similarly. However, as most Khasi participants in their experiment are women, the authors admit the possibility that Khasi female participants could have deduced the gender distribution of potential competitors. Accordingly, the authors acknowledge that their interpretation for the result [nurture] might be compromised if women are more likely to compete against women regardless of their society.

⁷ In an interesting study about how men's and women's competitive preference evolve early in life, Sutter and Rützler (2010) also include mixed- and single-sex groups in their design to investigate boys and girls.

the sex of potential competitors remains unclear. Moreover, a key difference between these studies and our experiment is that they inform participants about the gender composition in the lab before participants choose a payment scheme, which likely induces experimenter demand effects (e.g., Zizzo 2010).

Our study also speaks to recent research investigating institutional mechanisms to promote women to compete more. Motivated by the on-going debate about affirmative action policy, Balafoutas and Sutter (2012) and Niederle et al. (2013) find that instituting a competitive option more favorable to women in the form of quotas significantly increases women's entry into competition. As a result, the gender gap in competition entry decreases or even reverses. In addition, Balafoutas and Sutter (2012) show that preferential treatment for women in the form of a head start has a similar effect. However, they also find that preferential treatment in the form of a repetition of the competition in case a man wins—an institution that does not make the competition more sex-specific—does not significantly alter women's entry into competition.

Our findings provide evidence that women are not less competitive than men, but rather dislike competing against men. In line with the literature, we find that women's willingness to compete is significantly lower than men's when male and female participants are present in the lab, and there is no possibility of choosing the sex of a competitor. In stark contrast, we find that women's percentage of competition entry is high and similar to that of men when male and female participants are present in the lab, but they can choose the sex of the competitor while deciding to compete. Likewise, women's entry into competition is high when there is no possibility of choosing the sex of a competitor, but only female participants are present in the lab. Since the competitive payment scheme we offer in each condition is the same, collectively these results indicate that the significant gender gap in the replication condition is not due to women shying away from competition per se, but rather women shying away from competition against men. Corroborating this interpretation, we observe that women who compete in the choice of sex condition mostly choose a female competitor.

We find little support for negative self-stereotyping in our setting. Although we show that women hold the stereotype that men are better at mathematics, we find that differences in confidence between men and women do not explain much of the gender gap in the replication condition. Moreover, women's preference for a female competitor in the choice of sex condition is not explained by a higher confidence to compete against women than against men either. Notably, rather than observing significant confidence differences between men and women in the replication condition, we instead observe a significantly lower entry into competition among low-confident women compared to low-confident men. We reach the same conclusion when comparing low-confident women in the

However, the figure they report of girls' entry into competition is not broken down by single- and mixed-sex groups.

replication condition to low-confident women in the other two conditions. High-confident women, on the other hand, similarly enter competition across the three conditions. These results indicate that women's dislike of competition against men primarily affects the entry into competition of low-confident women.

The interpretation we give to women's behavior across conditions is substantiated by their beliefs about the sex of potential competitors. To directly test our conjecture that women's beliefs about the sex of potential competitors influence their willingness to compete, we elicit two kind of beliefs. First, we elicit if (and how) women perceive the gender composition of participants present in the lab. Second, we elicit women's beliefs about the likelihood of each sex to enter competition. We show that in each condition almost all women correctly perceive the actual gender composition of participants present in the lab. Moreover, in support of our hypothesis, we find in the replication condition that women are significantly less likely to enter competition the higher women believe is the proportion of male participants among the participants who choose competition.

Less intuitive is our finding that women's decision to enter competition is not driven by women's belief about men's likelihood to enter competition, but rather by their belief about the likelihood of other women to enter competition. More specifically, in each condition, we find that the more women believe other women compete, the more likely they are to enter competition. Yet, the average women's belief about other women's likelihood to enter competition is significantly lower in the replication condition, and this difference explains to a large extent the lower competition entry of women in the replication condition compared to women in the other two conditions.

Our experiment allows us to distil distinctive policy intervention to encourage women to enter competitive male-typed domains. First, our findings on women's beliefs about the sex of potential competitors offer a new perspective on how to *nudge* (see, e.g., Thaler and Sunstein 2008) women to compete. In light of our results, we prescribe that highlighting female role models can increase women's willingness to enter competitive male-typed domains. Namely, we propose that information in job advertisements that highlight women's underrepresentation—which is, for instance, a common practice in the economics academic job market⁸—should be replaced by references to recent female applicants and/or existing female workers. Second, our findings suggest that policy interventions that induce an increase in women's confidence can be effective, but not as plainly as we could expect. The evidence we present shows that women are not less confident than men while deciding to enter a mixed-sex competition. Rather, the gender difference we observe is that low-confident women are more reluctant to

⁸ Economics is a conspicuous male-typed domain because it is the least friendly science toward women among social sciences and the gender gap gets larger at each upper stage of the academic ladder. In 2014, only 29% of assistant professors, 23% of associate professors with tenure, and 12% of full professors were women (source: 2014 annual report of the Committee on the Status of Women in the Economics Profession, a committee of the American Economic Association).

compete compared to low-confident men. This implies that policy interventions should particularly target low-confident women.

The remainder of the chapter is organized as follows. In section 2.2, we describe the design of the experiment. In section 2.3, we present the results of the task performance, choice of payment scheme, and risk attitude. In section 2.4, we analyze participant's confidence level. In section 2.5, we assess participants' beliefs about the sex of potential competitors. In section 2.6, we discuss the policy implications of the results. In Section 2.7, we conclude.

2.2. Design of the experiment

To investigate men's and women's self-selection into competition, we conduct a laboratory experiment with three conditions: the *replication* condition, the *choice of sex* condition, and the *all women* condition.

2.2.1. Experimental procedure

The experiment is computerized using z-Tree software (Fischbacher 2007) and conducted in the Behavioral and Experimental Economics Laboratory (BEElab) at Maastricht University. We recruit the participants via ORSEE software (Greiner 2015) from the BEElab subject pool. In total, 204 subjects participate in the experiment (58, 90, and 56 in the replication condition, the choice of sex condition, and the all women condition, respectively), and predominantly have a business and economics background.⁹ By and large, participants also do not know each other at the session level.¹⁰ The average age is 22. Each session lasts approximately 60 minutes on average. Average earnings are € 12.65. Participants' decisions and interactions are kept anonymous. Moreover, we do not inform participants about the gender composition in the lab before eliciting their preference to compete in any of the conditions. However, importantly for our purposes, before being invited to take a seat in a private computer cubicle, participants spend a few minutes in a welcome room where they could see each other.

⁹ 78% of the participants have a business and economics background. The remaining participants have a background in law, health sciences, psychology, neuroscience, arts and culture, European studies, or computer science.

¹⁰ Using the debriefing question: "How many of the participants in this experimental session do you consider as a person you know?", we find that participants know, on average, 0.81, 0.71, and 0.64 persons in the replication, choice of sex, and all women conditions, respectively.

2.2.2. Experimental conditions

2.2.2.1 Replication condition

An equal number of men and women participate in each session of this condition.¹¹ We describe below the different steps of this condition in the order they happen.

- ***Practice round.*** The experiment starts with a practice round to familiarize each participant with the real-effort task. The task consists of consecutively solving five two-digit numbers addition problems (Niederle and Vesterlund 2007; e.g., $84 + 94 + 17 + 19 + 39$). Previous results indicate no significant gender difference in productivity to perform this addition task (e.g., Niederle and Vesterlund 2007). In this round, each participant practices the addition task for 2 minutes and receives no payment.¹²

- ***Round 1: Baseline Performance.*** Each participant performs the addition task for 5 minutes under a piece-rate payment scheme, earning € 0.50 per correct answer. This step serves to elicit the addition task productivity of each participant.

- ***Choice of payment scheme for a subsequent performance of the addition task.*** Each participant chooses between a non-competitive and a competitive payment scheme for a subsequent performance of the addition task. If a participant chooses the non-competitive alternative, the participant is paired with a randomly chosen participant among the participants who have also chosen the non-competitive alternative. By the end of the experiment, one of the two is chosen with a 50% probability for actual payment.¹³ The chosen participant earns € 1 per correct answer. The other participant earns nothing regardless of their number of correct answers. If a participant chooses the competitive alternative, the participant subsequently competes under a winner-take-all tournament against a randomly chosen participant among the participants who have also chosen to compete.¹⁴ In this pairwise competition, the participant who correctly solves more problems earns € 1 per correct answer, and the other earns zero. In case of a tie, one of the two is chosen as the winner with a 50% probability.

¹¹ We conduct two sessions in this condition. The procedure to ensure an equal number of male and female participants per session is described in the Appendix 2.A.3.

¹² During the practice round as well as during subsequent performance rounds, the following holds while participants perform the addition task: i. After an answer is provided, we immediately inform whether the answer is correct and a new problem is instantaneously displayed on the screen; ii. The total number of correct answers is permanently displayed on-screen; iii. We do not provide feedback on other participants' performance.

¹³ In case an odd number of participants choose the non-competitive payment, the unmatched participant is still paid with a 50% probability, but this participant being paid does not imply that another participant is not paid.

¹⁴ In case an odd number of participants choose competition, a randomly chosen performance among the matched participants who have chosen competition is used a second time to determine the earnings of the unmatched participant. If only one participant chooses competition (something that did not happen), the competition is not implemented and the participant is paid according to the non-competitive payment.

▪ ***Winning belief.*** Elicitation of each participant belief in winning the subsequent competition (we describe this elicitation in detail in section 2.4).

▪ ***Round 2: Performance under the chosen payment scheme.*** Each participant performs the addition task for 5 minutes under the chosen compensation scheme.

▪ ***Perception of the gender composition.*** Elicitation of each participant's perception of the gender composition of participants present in the lab (we describe this elicitation in detail in section 2.5).

▪ ***Beliefs about the inclination of each sex to enter competition.*** Elicitation of each participant's beliefs about the likelihood of each sex to enter competition (we describe this elicitation in detail in section 2.5).

▪ ***Stereotype-based beliefs.*** Elicitation of each participant's beliefs about men's and women's performance in Round 1, in which all participants perform under a piece-rate payment scheme (we describe this elicitation in detail in section 2.4).

▪ ***Risk attitude.*** Each participant answers the question: "How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, in which the value 0 means: 'not at all willing to take risks' and the value 10 means: 'very willing to take risks' ".¹⁵

▪ ***Questionnaires.*** Each participant answers demographics and debriefing questions.

We inform participants beforehand that one of the two monetarily incentivized performance rounds is randomly chosen with equal probability at the end of the experiment for actual payment. A participant's total payment in the experiment equals the sum of the earnings for the randomly chosen round, the earnings associated with monetarily incentivized belief elicitation questions plus a € 3 show up fee. Participants only learn of their earnings after completing the questionnaires.

Our measure for competitiveness builds on the existing literature. However, the non-competitive alternative we offer is different. The standard in the literature is to offer a plain piece-rate as a non-competitive payment alternative. A possible reason to observe choice differences between a plain piece-rate and a competitive payment scheme is differences in risk attitude because the latter incentive makes payment uncertain. Therefore, to mitigate the influence of risk attitude on the choice of payment scheme, we offer instead a non-competitive payment

¹⁵ In a field experiment, Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2011) show that responses to this general risk question reliably predict incentivized lottery choices.

alternative that also makes payment uncertain, although not dependent on the performance of other participants.

Another noteworthy aspect of the design is the implementation of a real-time competition.¹⁶ This competition setting allows us to inform each participant that choosing the competitive alternative means to subsequently compete against another participant who has also chosen to compete. As we argue in later sections, this competition setting is essential to elicit the proper confidence level of each participant as well as the beliefs of each participant about the sex of potential competitors.

2.2.2.2. Choice of sex condition

An equal number of men and women participate in each session of this condition.¹⁷ The only important difference compared to the replication condition is when participants choose a payment scheme for a subsequent performance of the addition task. In this step, each participant is rather asked to choose between three payment alternatives for a subsequent performance of the addition task. The first alternative is the same non-competitive payment scheme offered in the replication condition. The other two alternatives offer a competitive payment scheme. A participant who chooses the second alternative subsequently competes under a winner-take-all tournament against a randomly chosen *man* among the men who have also chosen to compete. A participant who chooses the third alternative subsequently competes under a winner-take-all tournament against a randomly chosen *woman* among the women who have also chosen to compete.¹⁸ The only difference between the two competitive alternatives is the sex of the competitor. We emphasize that the competitive payment scheme is the same; it is a winner-take-all tournament.

2.2.2.3. All women condition

We use the same design as in the replication condition. The distinctive element is that only women participate in this condition.¹⁹

¹⁶ The procedure routinely used in the literature to determine the earnings of a participant who chooses competition is to compare a participant's subsequent competitive performance to a past exogenously elicited competitive performance of other participants in the experiment. Yet, we are not the first to implement real-time competition (see Dohmen and Falk 2011).

¹⁷ We conduct three sessions in this condition. The procedure to ensure an equal number of male and female participants per session is described in the Appendix 2.A.3.

¹⁸ Participants who choose competition are matched to participants with equivalent choices (e.g., a man who chooses a female competitor is matched to a woman who chooses a male competitor). For unmatched participants, we use a similar procedure to the one described in footnote 14 (see Appendix 2.A.3 for the details of the procedure that we use in the choice of sex condition). For the cases in which the competition cannot be implemented (something that did not happen), the participant is paid according to the non-competitive payment.

¹⁹ We conduct two sessions in this condition.

2.3. Basic experimental results²⁰

2.3.1. Piece-rate baseline performance

We start by verifying that there is no significant gender difference in productivity for the addition task. In Table 2.1, we document the participants' piece-rate baseline performance. We observe that men's and women's performance is similar in the replication and the choice of sex conditions, respectively. Moreover, we reach the same conclusion after comparing the distribution of women's performance in the all women condition to the distribution of men's performance in the other two conditions (Mann-Whitney tests yield a p-value ≥ 0.542 for both comparisons). Therefore, the participants' baseline performance data do not corroborate the stereotype about women's lower ability in mathematics.²¹

**Table 2.1. Piece-rate baseline performance
(in average number of correct answers)**

Note: p-values derived from Mann-Whitney tests of equality of distributions between men and women. Standard deviation in parentheses.

	Replication	Choice of Sex	All Women
Men	10.31 (4.05)	10.29 (5.07)	n.a.
Women	10.28 (3.50)	10.38 (3.76)	10.45 (4.78)
p-value	0.919	0.418	n.a.

2.3.2. Choice of payment scheme

In Table 2.2, we summarize men's and women's self-selection into competition. In the replication condition 69% of men choose to compete whereas only 38% of women do so. This observed gender gap in competition entry is substantial and in line with the literature. In regression (1) of Table 2.3, we show that this result is robust when controlling for individual baseline performance and risk score. Specifically, women have a significant 32 percentage point lower probability than men of choosing the competitive payment scheme, *ceteris paribus*.

Table 2.2. Choice of a winner-take-all tournament (in percentage)

Note: p-values derived from Fisher's exact tests of contingency between men and women.

	Replication	Choice of Sex	All Women
Men	69%	76%	n.a.
Women	38%	67%	71%
p-value	0.034	0.486	n.a.

In the choice of sex condition, 76% of men and 67% of women select the competition. This gender gap in competition entry is not significant and robust when controlling for individual productivity and risk attitude (see regression (2) of Table 2.3). Recall that the two competitive alternatives in this condition involve a winner-take-all tournament, which is the same payment incentive that we offer as a competitive alternative in the replication condition. Hence, if women had a weaker preference to compete per se, we would have also observed a low percentage of competition entry among the women in the choice of sex condition. Instead, we observe a striking 76% increase in the percentage of women who choose competition in the choice of sex condition compared to the replication condition. In fact, the narrowing of the gender gap in entering competition in the choice of sex condition is due to a significant increase of women who choose to compete (38% vs. 67%, $p = 0.018$, Fisher's exact test) since the competition entry of men is only insignificantly higher in the choice of sex condition (69% vs. 76%, $p = 0.597$, Fisher's exact test).

Table 2.3. Probit models of payment choice

Note: The table reports marginal effects. *** significant at 1%. Robust standard errors in brackets.

Dependent variable: 1 if payment choice is a winner-take-all tournament			
	Replication (1)	Choice of Sex (2)	All Women (3)
<i>1 if female</i>	-0.32*** [0.123]	-0.11 [0.089]	n.a.
<i>Baseline performance</i>	0.05*** [0.016]	0.04*** [0.010]	0.04*** [0.010]
<i>Risk score</i>	-0.007 [0.028]	0.016 [0.017]	0.022 [0.025]
Observations	58	90	56
Pseudo R ²	0.164	0.120	0.133

As for the choice of the competitor's sex, we show in Table 2.4 that both men and women reveal preference for a female competitor, but this inclination is not significant for men. In addition, regression analysis shows that both men's and women's choice of the competitor's sex is not driven by their baseline performance and risk attitude (see Appendix 2.A.1).

Table 2.4. Choice of the competitor's sex (in percentage)

Note: p-values derived from binomial tests for the null hypothesis that the choice of a male or a female competitor is equally likely to occur.

	Male competitor	Female competitor	p-value
Men	35%	65%	0.121
Women	23%	77%	0.005

These results are consistent with our conjecture that the sex of potential competitors influences a woman’s decision to enter competition. To further investigate women’s behavior, we subsequently conduct the all women condition, in which the setting is the same as in the replication condition, but only female participants are present in the lab. In this condition, we find that 71% of women enter competition. This represents a striking 87% increase in the percentage of women who compete compared to the replication condition (38% vs. 71%, $p = 0.005$, Fisher’s exact test).²² Thus, women’s significantly higher percentage of competition entry in the all women condition is also consistent with our hypothesis.

Finally, a remark on risk attitude is in order. In regressions (1) – (3) of Table 2.3 the risk score predictor is statistically insignificant in each condition, which indicates that participants’ choice of payment scheme is not determined by their risk attitude.²³ This is likely due to the design of our experiment in which payment is risky under both the competitive and non-competitive payment schemes.

2.4. Stereotypes, confidence, and the “irrational” competitors

In this section, we investigate whether there is evidence to establish a connection between the stereotype that men are better at mathematics and the choice for competition. Specifically, we investigate whether we can discriminate between two possible explanations grounded in stereotype threat: confidence to compete vs. preference to compete. A common element to both explanations is that each assumes that individuals of the targeted group hold the stereotype that men are better at mathematics. Therefore, we start by verifying that participants do indeed hold the stereotype.

2.4.1. Stereotype-based beliefs

To verify that participants hold the stereotype we elicit participants’ estimation of the performance of male and female participants in the 5 minutes Round 1 piece-rate performance. Specifically, each participant answers two questions: i. “How many addition problems did male participants correctly solve, on average, in the 5 minutes Round 1 piece-rate performance (excluding yourself if you are a man)?”; ii. The same question, but with reference to female participants. These questions are monetarily incentivized using a scoring rule that yields a higher reward the more accurate the estimate is (see Appendix 2.A.4 for the full instructions, which provide the details of the monetary incentive scheme).²⁴

²² Compared to the choice of sex condition, the increase in the percentage of competing women is not significant (67% vs. 71%, $p = 0.667$, Fisher’s exact test).

²³ The risk score remains insignificant when we regress specification (1) and (2) of Table 2.3 separately for men and women.

²⁴ In the all women condition, we ask participants the same two questions, but about male and female participants of a previous session of the experiment so that we avoid deception.

In the first row of Table 2.5, we report the average of men’s estimate of the gender gap in performance per condition. In the second row of Table 2.5, we report the corresponding figures for women. We find that participants hold the stereotype in each condition. That is, participants do believe that men have, on average, a significant importantly, the participants of the targeted group—women—hold the stereotype.

**Table 2.5. Participants’ estimate of the gender gap in performance
(in number of correct answers)**

Note: A participant’s estimate of the gender gap equals a participant’s estimate of the performance for men minus a participant’s estimate of the performance for women. *** significant at 1% derived from Wilcoxon signed-rank tests. Standard deviation parentheses.

	Replication	Choice of Sex	All Women
Men	1.1*** (2.70)	0.8*** (2.37)	n.a.
Women	2.1*** (2.52)	1.7*** (2.22)	1.6*** (3.06)

Hence, women’s competition entry behavior across conditions accommodates a preference-based explanation grounded in stereotype threat. As we discussed in the introduction, if women are chronically exposed to a math stereotype threat relatively to men, a logical reaction of women is to develop a preference to avoid situations where their math ability can be compared to men. Moreover, women’s dislike with being compared to men is amplified in our setting because we include a competitive dimension. In light of this explanation, women’s entry into competition should be low in the replication condition, and they should prefer a female competitor in the choice of sex condition. In the all women condition, there is no stereotype threat (assuming women perceive that men are not present). As a result, women’s entry into competition should be high.

However, as the stereotype pertains to women’s lower ability in mathematics, an alternative mechanism could be that women’s exposure to a situation evoking the stereotype triggers negative self-stereotyping, which undermines women’s confidence to compete against men. Accordingly, women’s belief in winning the competition should be considerably lower in the replication condition because women cannot avoid the possibility of a mixed-sex competition when they choose competition. In contrast, women’s belief in winning the competition is not biased downwards in the choice of sex condition because negative self-stereotyping bias women’s confidence against men, but not against other women. In the all women condition, negative self-stereotyping does not come to mind at all (assuming women perceive the absence of male participants). Therefore, according to the self-stereotyping model, the lower percentage of competition entry among women

in the replication condition is a rational response of women as payoff maximizers to their belief in a lower probability of winning the competition.

To test self-stereotyping, the stereotype-based beliefs that we report above are not appropriate. The stereotype-based beliefs surely indicate how a female participant compares the addition task performance of other female participants relative to male participants. However, these beliefs do not necessarily indicate how a female participant believes in her *own* performance of the addition task relative to the *other participants who choose to compete*, which is in our view the proper measure of a participant’s confidence level at the moment they choose a payment scheme.

2.4.2. Winning belief: the impact of confidence on the choice of payment scheme

To sort out the impact of possible differences in confidence level on the choice of payment scheme, we elicit participants’ belief in winning the competition.

We elicit this belief immediately after participants make their choice of payment scheme. In the replication and all women conditions, we ask participants the following question: “Consider only the other participants who have just chosen to compete in the subsequent performance round. How many of these participants (in percentage terms) did correctly solve less addition problems than you in the 5 minutes Round 1 piece-rate performance?”. In the choice of sex condition, we ask participants two versions of this question. In one version, we ask the question with reference to the male participants who have chosen to compete; in the other version, we ask the question with reference to the female participants who have chosen to compete. The questions are monetarily incentivized using a scoring rule that yields a higher reward the more accurate the estimate is (see Appendix 2.A.4 for the full instructions, which provide the details of the monetary incentive scheme).

Ideally, the winning belief should incorporate two elements: i. Elicit a participant’s belief about their *own* performance relative to *only* the *other participants who have chosen to compete*; ii. A participant’s estimate is about their expected relative performance in the *subsequent competition*. We only incorporate the former element because incorporating the latter element poses a serious incentive problem.²⁵ As a proxy for the second element, we use a participant’s estimate of their relative performance in the piece-rate round. That is, we rely on the assumption that a participant’s belief about their relative performance among the

²⁵ If the question were instead “Consider only the other participants who have just chosen to compete in the subsequent performance round. How many of these participants (in percentage terms) do you expect to correctly solve less addition problems than you in the subsequent performance round?”, a too salient hedging opportunity would arise because we monetarily incentivize the question using a scoring rule that makes a participant’s earnings higher the more accurate a participant’s estimate is. The scoring rule would, therefore, give an incentive for a participant to report an estimate of zero and subsequently solve no problem to ensure the maximum payment.

participants who choose competition is similar in both rounds. Spearman rank correlations are consistent with this assumption [replication condition: piece-rate performance (PR) vs. competitive performance (COM), $\rho = 0.731$. Choice of sex condition: PR Men vs. COM Men, $\rho = 0.749$ | PR Women vs. COM Women, $\rho = 0.775$. All women condition: PR vs. COM, $\rho = 0.847$].

In Table 2.6, we document the winning beliefs. If the self-stereotyping hypothesis were correct, the women’s belief in winning in the replication condition—in which they cannot avoid the possibility of competing against men—should be substantially lower not only than men’s belief in winning but also than women’s belief in winning in both the choice of sex and all women conditions. However, none of these comparisons leads to significant differences (41.5% vs. 44.5%, $p = 0.791$, Mann-Whitney test; 41.5% vs. 44.9%, $p = 0.527$, Mann-Whitney test; 41.5% vs. 44.6%, $p = 0.556$, Mann-Whitney test). Although women in the replication condition are directionally less confident, the difference in confidence level is not significant. In other words, we do not find evidence for negative self-stereotyping.

Table 2.6. Winning belief (in average percentage)

Note: In the choice of sex condition, a participant’s winning belief equals $\max\{\text{winning belief against a man, winning belief against a woman}\}$. Standard deviation in parentheses.

	Replication	Choice of Sex	All Women
Men	44.5% (24.9)	44% (25.6)	n.a.
Women	41.5% (24.4)	44.9% (25.7)	44.6% (24.7)

Regression analysis corroborates this finding. Comparing regressions (1) and (4) of Table 2.7, we see that after adding the winning belief as a predictor, the gender gap in the probability of entering competition (which is measured by the female predictor) in the replication condition only slightly decreases. In addition, comparing regressions (2) and (5) of Table 2.7, we see that the gender gap in competition entry in the choice of sex condition remains insignificant after including the winning belief as a predictor. In short, although regressions (4) – (6) of Table 2.7 show that the winning belief helps to predict entry into competition in each condition, differences in men’s and women’s confidence level do not explain much of the observed gender difference in entry into competition in the replication condition.²⁶

²⁶ In stark contrast to the winning beliefs, we find no significant effect of the stereotype-based beliefs when we add them as a covariate in the regressions of Table 2.7 (as well as in the remaining regression analyses that we present in this chapter). Thus, we dismiss this variable in the regression analyses that we present for the sake of relevance.

Table 2.7. Probit models of payment choice II

Note: The table reports marginal effects. In Panel A, we again report the results of Table 2.3 for ease of comparison. *Winning belief* is a variable that ranges from 0-100. ***, **, and * significant at 1%, 5%, and 10%, respectively. Robust standard errors in brackets.

Dependent variable: <i>I if payment choice is a winner-take-all tournament</i>						
	Panel A			Panel B		
	Replication (1)	Choice of Sex (2)	All Women (3)	Replication (4)	Choice of Sex (5)	All Women (6)
<i>I if female</i>	-0.32*** [0.123]	-0.11 [0.089]	n.a.	-0.28*** [0.106]	-0.11 [0.084]	n.a.
<i>Baseline performance</i>	0.05*** [0.016]	0.04*** [0.010]	0.04*** [0.010]	0.03** [0.017]	0.02* [0.012]	0.02 [0.013]
<i>Risk score</i>	-0.007 [0.028]	0.016 [0.017]	0.022 [0.025]	-0.004 [0.023]	0.017 [0.017]	0.024 [0.023]
<i>Winning belief</i>				0.009*** [0.002]	0.006*** [0.002]	0.005** [0.002]
Observations	58	90	56	58	90	56
Pseudo R ²	0.169	0.120	0.133	0.388	0.197	0.190

Table 2.8. Choice of a winner-take-all tournament split by low- and high-confident participants

Note: *bwin* refers to participants' winning belief. p-values derived from Fisher's exact tests of contingency between men and women. The choices of the participants with a winning belief equal to 50% are: in the replication condition, one woman competes and two women do not compete; in the choice of sex condition, one woman and one man compete, and one woman does not compete; in the all women condition, three women compete.

	Replication		Choice of Sex		All Women	
	<i>bwin</i> < 50	<i>bwin</i> > 50	<i>bwin</i> < 50	<i>bwin</i> > 50	<i>bwin</i> < 50	<i>bwin</i> > 50
Men	56%	91%	62%	95%	n.a.	n.a.
Women	17%	88%	56%	84%	61%	82%
p-value	0.035	1.000	0.779	0.603	n.a	n.a

Nonetheless, we find an interesting difference related to the confidence level. In Table 2.8, we break down the entry into competition by participants with a winning belief lower than 50%—who we denominate as low-confident—and participants with a winning belief higher than 50%—who we denominate as high-confident. In the replication condition, we observe a considerable lower entry into competition among low-confident women compared to low-confident men. The same is observed when comparing low-confident women in the replication condition to low-confident women in the other two conditions. High-confident women, on the other hand, similarly enter competition across the three conditions. Collectively, these findings show that the lower inclination of women to enter competition in the replication condition is driven by the low-confident women.

2.4.3. Confidence and the choice of sex

The results so far show that women shy away from competition only when they cannot avoid the possibility of competing against men. Since this behavior is not explained by women's lower confidence level relatively to men, we understand the results as evidence of women's dislike of competition against men. However, women's aversion to competition in the replication condition seems to hold only for low-confident women. As a robustness check for whether women's dislike of competition against men is specific to low-confident women, we examine the impact of confidence level on women's choice of the competitor's sex.

In Table 2.9, we show that both low- and high-confident women overwhelmingly, and equally, prefer to compete against a woman. Moreover, regression analysis shows that women's choice of the competitor's sex is not driven by a higher confidence level to compete against a woman (see Appendix 2.A.1 for details). Looking at men, we observe that both low- and high-confident men's choice of the competitor's sex is not significantly different from a random choice.

Table 2.9. Choice of the competitor's sex split by low- and high-confident participants

Note: *bwin* refers to participants' winning belief. p-values derived from binomial tests for the null hypothesis that the choice of a male or a female competitor is equally likely to occur.

		Male competitor	Female competitor	p-value
Men	<i>bwin</i> > 50	35%	65%	0.332
	<i>bwin</i> < 50	38%	63%	0.454
Women	<i>bwin</i> > 50	20%	80%	0.035
	<i>bwin</i> < 50	21%	79%	0.057

Our conclusion is that women dislike competing against men independently of their confidence level. However, if women cannot avoid the possibility of competing against men when women enter competition (as is the case in the replication condition), high-confident women still enter competition because their belief in a high probability of winning the competition more than compensates their dislike of competition against men.

2.4.4. The “irrational” competitors

In this section, we examine whether a participant’s choice of payment scheme is a rational response to a participant’s winning belief in terms of expected payoff maximization.

Under reasonable assumptions, we can predict the choice of payment scheme that is consistent with expected payoff maximization solely based on a participant’s winning belief (see the Appendix 2.A.2 for the detailed theoretical analysis). In a nutshell, assuming a participant is risk neutral or risk averse²⁷, a participant who believes to have a higher (lower) than 50% chance of winning the competition should choose the winner-take-all tournament (non-competitive payment).

In the following analysis, we compare each participant actual choice of payment scheme to the choice predicted once considering their winning belief. Logically, a participant has a rational response if a participant’s actual choice of payment scheme coincides with the predicted choice.

In column (3) of Table 2.10, we report the percentage of participants whose actual choice coincides with the predicted choice. We observe that 86% of women in the replication condition have a rational response to their winning belief. The percentage of rational responses in the other four cases—men in the replication condition, men and women in the choice of sex condition, and women in the all women condition—is significantly lower compared to women in the replication condition (Fisher’s exact tests of contingency yield $p\text{-value} \leq 0.07$ for any of the four comparisons). The lower percentage of rational responses in these four cases seems to be due to an excessive entry into competition. If all participants had a rational response, then only 38% of men in the replication condition, 41% of men and 42% of women in the choice of sex condition, and 42% of women in the all women condition would have chosen competition (see column (1) in Table 2.10).

To examine excess entry, we analyze the choices of the participants who do not have a rational response. There are two possible types of “irrational” response in our setting: i. The “*irrational*” competitor, i.e., a participant who chooses competition, but whose expected payoff maximization choice corresponding to their winning

²⁷ It is reasonable to assume that participants are either risk neutral or risk averse since economics experiments eliciting risk attitude do not find support for risk loving preferences (see Croson and Gneezy 2009).

Table 2.10. Rational responses in terms of expected payoff maximization

Note: **Predicted Choice** refers to the predicted percentage of competitive choices given participants' winning belief; **Rational Response** refers to the percentage of participants whose actual choice coincides with the predicted choice. In column (1), we treat the few participants who estimate a probability of winning equal to 50% as equally likely to choose between the two payment alternatives. We again report the percentage of actual competitive choices for ease of comparison.

		Predicted Choice (1)	Actual Choice (2)	Rational Response (3)
Replication	Men	38%	69%	62%
	Women	33%	38%	86%
Choice of Sex	Men	41%	76%	62%
	Women	42%	67%	62%
All Women	Women	42%	71%	59%

belief is the non-competitive payment; ii. The “*irrational*” *non-competitor*, i.e., a participant who chooses not to compete, but whose expected payoff maximization choice corresponding to their winning belief is competition.

The results in Table 2.11 show that the occurrence of both types of “irrational” response is sparse among women in the replication condition. In the other four cases, however, we observe that the percentage of “irrational” competitors is considerable and clearly higher than the percentage of “irrational” non-competitors. McNemar tests for the significance of changes validate these observations. The occurrence of both responses is not significantly different for women in the replication condition whereas in each of the other four cases the “irrational” competitor response occurs significantly more often (McNemar tests yield $p\text{-value} = 0.625$ for women in the replication condition and $p\text{-value} \leq 0.013$

Table 2.11. “Irrational” competitors and “irrational” non-competitors

Note: “Irrational” competitors (non-competitors) refers to the participants who choose competition (no competition), but whose choice consistent with expected payoff maximization would be the non-competitive payment (competitive payment).

	Replication		Choice of Sex		All Women	
	“Irrational” competitors	“Irrational” non-competitors	“Irrational” competitors	“Irrational” non-competitors	“Irrational” competitors	“Irrational” non-competitors
Men	35%	3%	36%	2%	n.a.	n.a.
Women	10%	3%	31%	7%	34%	7%

for any of the other four cases).²⁸

In short, we observe that in all cases, except for women in the replication condition, the percentage of “irrational” competitors is substantial. We regard an “irrational” competitor response as stronger evidence of a participant’s preference to compete than if a participant’s choice to compete is in accordance with expected payoff maximization.²⁹ Hence, we argue that the observed low percentage of “irrational” competitors only among women in the replication condition is evidence that women dislike competing against men.

2.5. Uncovering the beliefs about the sex of potential competitors

We conclude our analysis by examining participants’ beliefs about the sex of potential competitors. Since participants are not informed about the gender composition of participants present in the lab, we first examine if (and how) each participant perceives the gender composition of participants present in the lab. Second, we elicit participants’ beliefs about the likelihood of each sex to enter competition. Finally, we analyze whether the beliefs about the sex of potential competitors influence participants’ decision to enter competition.

2.5.1. Perception of the gender composition

To assess if (and how) participants perceive the gender composition of participants present in the lab we ask each participant the following question: “How do you perceive the gender composition of participants in this experimental session?”. The response options available are “I did not notice the gender composition of participants”; “Only female participants”; “Mainly female participants”; “Balanced composition”; “Mainly male participants”; and “Only male participants”. This elicitation is not monetarily incentivized with a scoring rule because rewarding an accurate perception would have incentivized the participants who did not notice on the gender composition to guess it.

The results in Table 2.12 indicate that the overwhelming majority of women reveal that they did notice the gender composition whereas this aspect has not distinctively attracted the attention of men. Hence, we proceed the analysis considering only women for the sake of relevance and brevity (see Appendix 2.A.1 for the corresponding analysis for men).

²⁸ The McNemar test analyzes whether the two possible types of “irrational” responses happen with the same frequency. The null hypothesis for each of the 5 cases we test is that the proportion of “irrational” competitors and “irrational” non-competitors is the same. This test is not sensitive to the number of rational competitors and rational non-competitors.

²⁹ Reuben, Wiswall, and Zafar (2013) share a similar view.

Table 2.12. Unawareness of the gender composition (in percentage)

Note: The percentage refers to the men (women) who chose the alternative “I did not notice the gender composition of participants”.

	Replication	Choice of Sex	All Women
Men	45%	53%	n.a.
Women	14%	13%	16%

In Table 2.13, we report the specific gender composition of participants present in the lab that was perceived by the women who reveal to have noticed the gender composition. We see that in each condition almost every woman correctly perceives the actual gender composition of participants present in the lab. This finding supports that the beliefs we present in the following section, which are elicited based on the actual gender composition of participants present in the lab, capture the true beliefs of women at the moment they are asked to choose a payment scheme.

Table 2.13. Women’s perception of the gender composition (in percentage)

Note: In each condition, the percentages are computed relatively to the women who reveal to have noticed the gender composition.

	Only female	Mainly female	Balanced composition	Mainly male	Only male
Replication	0%	4%	88%	8%	0%
Choice of Sex	0%	8%	90%	2%	0%
All Women	87%	11%	0%	2%	0%

2.5.2. Beliefs about the likelihood of each sex to enter competition

To elicit participants’ beliefs about the sex of potential competitors, we ask participants the following two questions (presented on the same screen): i. “For your information, there are 15 male participants in this experimental session. How many male participants do you think have chosen to compete (excluding yourself if you are a male participant and chose competition)?”; ii. The same question, but with reference to female participants. The questions are monetarily incentivized using a scoring rule that yields a higher reward the more accurate the estimate is (see Appendix 2.A.4 for the full instructions, which provide the details of the monetary incentive scheme). In the all women condition, we only ask the second question, which we adjust for the higher number of female participants present in the lab.

In the subsequent analysis, we include all the female participants. The results are qualitatively the same if we exclude the few female participants who reveal not to have noticed the gender composition. In Table 2.14, we report women's beliefs about the likelihood of each sex to enter competition. We observe that women's belief about men's likelihood to enter competition is essentially the same in the replication and the choice of sex condition (70.6% vs. 72%, $p = 0.806$, Mann-Whitney test). Women's belief about other women's likelihood to enter competition is similar in the choice of sex and all women conditions (64.4% vs. 65.4%, $p = 0.897$, Mann-Whitney test), but significantly lower in the replication condition (51.5% vs. 64.4%, $p < 0.01$, Mann-Whitney test; 51.5% vs. 65.4%, $p < 0.01$, Mann-Whitney test). In other words, in the condition in which women cannot avoid the possibility of facing a male opponent when they compete, women's belief about other women's willingness to compete is, on average, substantially lower.

Table 2.14. Women's beliefs about the likelihood of each sex to enter competition

Note: **Belief probability men enter** equals: (estimate of male participants who enter competition/total male participants in the session)*100; **Belief probability women enter** equals: [estimate of female participants who enter competition/(total female participants in the session-1)]*100. We report the average of these variables. Standard deviation in parentheses.

	Replication	Choice of Sex	All Women
Belief probability men enter	70.6% (15.1)	72% (15.9)	n.a.
Belief probability women enter	51.5% (18.0)	64.4% (15.9)	65.4% (16.2)

2.5.3. Women's beliefs about the sex of potential competitors and competition entry

In regressions (1) and (2) of Table 2.15, we observe that women's belief about the percentage of men among the participants who choose competition significantly predicts women's choice for competition in the replication condition, but not in the choice of sex condition. The insignificance of this predictor in the choice of sex condition is logical because the proportion of men among the participants who choose competition does not affect the chance of facing a male competitor when a participant chooses to compete. In contrast, this predictor captures women's belief in the probability of facing a male competitor in the replication condition. Therefore, the significance of this predictor in the replication condition corroborates the finding that women dislike competing against men. More specifically, regression (1) of Table 2.15 indicates that women who believe in a higher probability of facing a male competitor are significantly less likely to choose competition, when controlling for performance, risk score, and winning belief.

Table 2.15. Probit models of payment choice III (only women)

Note: The table reports marginal effects. *Belief percentage men among competitors* indicates the women's belief about the proportion of men among the participants who choose competition, i.e., it equals: [estimate of male participants who enter competition / (estimate of male participants who enter competition + estimate of female participants who enter competition)] * 100. *Winning belief* is a variable that ranges from 0-100. *** and ** significant at 1% and 5%, respectively. Robust standard errors in brackets.

Dependent variable: 1 if payment choice is a winner-take-all tournament		
	Replication (1)	Choice of sex (2)
<i>Baseline performance</i>	0.03 [0.040]	0.02 [0.021]
<i>Risk score</i>	-0.02 [0.043]	0.01 [0.026]
<i>Winning belief</i>	0.010*** [0.002]	0.005** [0.002]
<i>Belief percentage men among competitors</i>	-0.024** [0.012]	-0.014 [0.009]
Observations	29	45
Pseudo R ²	0.468	0.191

Less intuitive is the more detailed result we present in regression (4) of Table 2.16, in which we separately report the impact of women's beliefs about men's and other women's likelihood to enter competition. We see that women's lower inclination to compete in the replication condition is not driven by women's belief about men's likelihood to compete, but rather by women's belief about other women's likelihood to enter competition. Moreover, regressions (5) and (6) of Table 2.16 show that in the choice of sex and all women conditions the more a woman believes other women enter competition, the more likely a woman enters competition herself. Finally, relevant for the policy implications that we draw in the next section, comparisons of regressions (4) – (6) with regressions (1) – (3) in Table 2.16 further show that the women's belief about the likelihood of other women to enter competition and the women's winning belief predictors are orthogonal.

In sum, in each condition, women's willingness to compete is higher the more they believe other women do compete. However, women's belief about other women's likelihood to enter competition in the replication condition is, on average, significantly lower. Put together, these findings indicate that women compete less in the replication condition because they believe fewer women compete.

The specification we present in Table 2.17, in which we pool the three conditions, clearly demonstrates the latter connection. Comparing regressions (1) – (3) of Table 2.17, we observe that adding women's belief about the likelihood of other women to enter competition as a predictor to the set of controls makes the

Table 2.16. Probit models of payment choice IV (only women)

Note: The table reports marginal effects. *Winning belief*, *Belief probability men enter*, and *Belief probability women enter* are variables that range from 0-100. *** and ** significant at 1% and 5%, respectively. Robust standard errors in brackets.

Dependent variable: <i>1 if payment choice is a winner-take-all tournament</i>						
	Panel A			Panel B		
	Replication (1)	Choice of Sex (2)	All Women (3)	Replication (4)	Choice of Sex (5)	All Women (6)
<i>Baseline performance</i>	0.03 [0.033]	0.03 [0.021]	0.02 [0.013]	0.03 [0.038]	0.03 [0.016]	0.01 [0.011]
<i>Risk score</i>	0.01 [0.035]	0.01 [0.026]	0.02 [0.023]	-0.01 [0.041]	-0.03 [0.025]	0.01 [0.024]
<i>Winning belief</i>	0.010*** [0.002]	0.005** [0.003]	0.005** [0.002]	0.010*** [0.002]	0.005** [0.002]	0.005** [0.002]
<i>Belief probability men enter</i>				-0.008 [0.007]	0.005 [0.004]	n.a.
<i>Belief probability women enter</i>				0.011** [0.005]	0.014*** [0.003]	0.010*** [0.003]
Observations	29	45	56	29	45	56
Pseudo R ²	0.340	0.166	0.190	0.492	0.366	0.322

Table 2.17. Probit models of payment choice V (only women, pooling the three conditions)

Note: The table reports marginal effects. Women in the choice of sex condition are the base group. *Winning belief* and *Belief probability women enter* are variables that range from 0-100. *** and * significant at 1% and 10%, respectively. Robust standard errors

Dependent variable: <i>1 if payment choice is a winner-take-all tournament</i>		
	Replication (1)	Choice of sex (2)
<i>Baseline performance</i>	0.03 [0.040]	0.02 [0.021]
<i>Risk score</i>	-0.02 [0.043]	0.01 [0.026]
<i>Winning belief</i>	0.010*** [0.002]	0.005** [0.002]
<i>Belief percentage men among competitors</i>	-0.024** [0.012]	-0.014 [0.009]
Observations	29	45
Pseudo R ²	0.468	0.191

difference in self-selection into competition between women in the replication condition and women in the choice of sex condition statistically insignificant. Moreover, we obtain the same conclusion when we regress specifications (1) – (3) of Table 2.17 using either women in the all women condition as the base group or women in the choice of sex and all women conditions pooled as the base group. Put simply, differences in women’s belief about other women willingness to compete largely explain the observed lower inclination of women to enter competition in the replication condition compared to women in the choice of sex and all women conditions

2.6. Discussion of policy implications

Our results indicate that implementing single-sex competition, or allowing women to choose the sex of their competitors promotes women to compete more. Such institutional changes, however, are usually out of the question.³⁰ Nevertheless, understanding the competitive behavior of women in the single-sex and the choice of sex contexts provides insight into interventions to encourage women to compete in the more realistic context in which they cannot avoid the possibility of competing against men.

2.6.1. Increase women’s willingness to compete via female role models

The evidence we present shows a systematic inclination of women to enter competition the more they believe other women may enter. This finding reveals women’s preference for a more sex-specific competition. One intervention in mixed-sex competitions that likely induces women to perceive the competition more sex-specific is affirmative action. Implementing a competition in mixed-sex groups with two possible winners, Niederle et al. (2013) find that women increase their competition entry from 31% to 83% when they introduce a quota ensuring that the better performing woman is among the winners. However, this quota intervention also illustrates the likely side effect that would emerge if such a policy were implemented outside the lab; men decrease their competition entry from 74% to 45% when the female quota is introduced.³¹

Behavioral economics research shows that choices are influenced by (according to standard rational choice theory) irrelevant features of the choice environment such as framing (e.g., Kahneman and Tversky 2000). Moreover, the ongoing *nudge* movement (see, e.g., Thaler and Sunstein 2008) advocates we can *nudge* better decisions without constraining freedom of choice if we properly design the choice

³⁰ Notable exceptions exist, though. In chess, for instance, women may choose to compete in mixed-sex or only-women tournaments.

³¹ From an efficiency perspective, we could argue that discouraging men is good if it drives out the less productive men. Yet, we leave aside that perspective in our analysis. Our experimental setting is not appropriate to quantify the cost/gain associated to participants’ entry into competition based on participants’ work task productivity because we purposely align the payoffs of the payment alternatives that we offer.

environment.³² Without wanting to suggest that the use of a quota system cannot be justified in any circumstance, we propose an alternative intervention to promote women's entry into male-typed domains: drawing women's attention to female workers in male-typed domains could also persuade women to enter these domains. To illustrate this idea, consider the academic job market in economics. In this male-typed domain, job advertisements commonly highlight women's underrepresentation as a form to encourage women (see, e.g., <https://econjobmarket.org/>). According to our results, however, this practice should be absolutely avoided. Rather than highlighting that women are a minority, job advertisements should instead include positive references to recent female applicants and/or existing female workers.³³

In brief, we assert that including information that highlights female role models rather than on-going interventions that highlight women's underrepresentation is the appropriate way to boost the participation of women in male-typed domains. Moreover, the *nudge* that we advocate could be an excellent alternative to controversial policies such as affirmative action in the form of quotas because the intervention that we propose would be unlikely to affect the behavior of men.

2.6.2. Increase women's willingness to compete via confidence level

In our experimental setting, we observe that higher confidence significantly predicts more entry of women into competition in every condition. Moreover, the predictive magnitude of women's confidence level is stronger in the replication condition.³⁴

Hence, our findings indicate that a policy intervention that increases the confidence level of women should boost their entry into competition, and this kind of intervention should be particularly effective in situations where women cannot avoid the possibility of competing against men. Furthermore, we also find that women on the verge of a mixed-sex competition are not less confident than men, but rather low-confident women enter much less into competition compared to low-confident men. Thus, interventions that boost women's confidence level should particularly target low-confident women because they are more susceptible to modify their competitive behavior.

³² Policy makers are increasingly open to the idea of nudging people as evident from the many examples on display at the [Nudge blog](#), and the establishment of the *Behavioural Insights Team* and *Social and Behavioral Sciences Team* by the UK and US governments, respectively.

³³ For instance, we suggest that a framing such as “The organization # wants to increase its proportion of female staff in areas where women are underrepresented. Women are therefore explicitly encouraged to apply” (source: <https://econjobmarket.org/>) should be replaced by a statement along the lines of “The recent considerable increase of our female employees has been contributing to our aim to promote a balanced representation of men and women among our staff. Therefore, we strongly encourage women to keep applying”.

³⁴ The significance of women's winning belief as well as the explanatory power of the specification after adding the winning belief as a predictor is stronger in the replication condition. Regressing Panel A models of Table 2.16 without the winning belief predictor yield a pseudo R^2 of 0.105, 0.116 and 0.133 in the replication, choice of sex, and all women conditions, respectively.

Our results may thus explain why a less invasive intervention, such as advice for women, is ineffective (see Brandts, Groenert, and Rott 2014)³⁵, but a strong intervention, such as preferential treatment for women in the form of a head start, boosts women's competition entry in a mixed-sex context (see Balafoutas and Sutter 2012). However, the results of Balafoutas and Sutter (2012) also illustrate the likely side effect that would emerge if such a policy were implemented outside the lab; when comparing Balafoutas and Sutter's (2012) major head start treatment to their treatment without intervention, women's competition entry increases from 31% to 69%, but men's competition entry decreases from 64% to 50%.³⁶

In general, there is room to intervene via women's confidence level. Our results particularly suggest that boosting the confidence of low-confident women is critical to making women compete more in a mixed-sex context. However, more research is needed to test specific interventions that could encourage women to compete without discouraging men. Last but not least, our results also show that confidence level positively influences women's inclination to compete independently of their belief about other women's willingness to compete, which indicates that the two types of intervention we discussed are complementary.

2.7. Concluding remarks

The present study was designed to determine whether different cues to the sex of potential competitors alter women's willingness to compete in a male-typed domain. To the best of our knowledge, our study represents the first attempt at explicitly testing in the lab the connection between women's beliefs about the sex of potential competitors and their decision to enter competition. Our hypothesis was that women do not have a weaker preference to compete per se, but rather dislike competing against men.

Our findings shed new light on the literature on gender competition. Our results demonstrate that women dislike competing against men in a male-typed domain. Specifically, we find that women's unwillingness to enter a potential mixed-sex competition is significantly driven by women's belief in a higher proportion of men among potential competitors. In contrast, if women can avoid the possibility of a mixed-sex competition when they enter competition, women compete as much as men do.

Moreover, we find the intriguing result that women's lower inclination to enter a mixed-sex competition in a male-typed domain is not driven by their belief about men's likelihood to enter competition, but rather by women's belief that

³⁵ Concerned with the reverse discrimination costs that affirmative action policy could entail, Brandts et al. (2014) investigate whether advice by more experienced individuals could be an alternative to promote women to compete more. They find that this intervention does not significantly increase women's competition entry.

³⁶ In unreported regressions, in which we run the models (1) and (2) of Table 2.16 for men, we find that higher confidence significantly predicts more entry into competition. Thus, lower confidence may explain men's lower entry into competition when there is a head start for women.

other women are less inclined to enter competition. It is the latter belief that in turn increases women's belief in facing a male competitor. Perhaps women behave according to what they believe other women do just because they are a woman themselves. This would be consistent with the psychology literature on gender typing (e.g. Bem 1981) and the sociology literature on homophily (e.g., McPherson, Smith-Lovin, and Cook 2001), which is the tendency of people to interact with individuals who resemble them. Future research will have to test the relevance of these motives.

We confirm previous findings that women's confidence level predicts their willingness to enter competition. However, we find little support for negative self-stereotyping (Bordalo et al. 2014), which in our setting would imply a lower confidence level of women when they cannot avoid the possibility of a mixed-sex competition. In this situation, we instead observe a remarkable lower inclination to enter competition of low-confident women compared to low-confident men.

Our results have important policy implications. A growing literature that investigates institutional mechanisms to promote women to compete more (e.g., Balafoutas and Sutter 2012; Niederle et al. 2013) shows that preferential treatment in the form of quotas or a head start for women boosts women's competition entry (but also decreases men's competition entry). Our findings suggest an alternative course of action to *nudge* women to compete that is unlikely to affect men's willingness to compete and, consequently, to be less controversial. While attempting to *nudge* women to enter competitive male-type domains, we should make reference to women who do enter or seek to enter these domains, rather than highlighting women's underrepresentation. Namely, including information in job advertisements about women who recently applied and/or got hired in male-typed domains deserves consideration as a promising intervention to explore.

In closing, some words of caution are in order. First, in contrast to the literature, our results are derived in a setting in which women's risk attitude does not influence their decision to enter competition. Therefore, whether our findings are transferable to a setting in which risk attitude is important is an open question that our data cannot answer.

Second, some studies quantify the cost/gain associated with participants' entry into competition based on participants' productivity of a real-effort task (see, e.g., Niederle et al. 2013, Balafoutas and Sutter 2012). Our setting, though, is unsuitable to perform such kind of analysis because we purposely align the payoffs of the payment alternatives we offer.

Third, our study only considers the supply-side. Even if supply-oriented interventions are effective in increasing women's willingness to enter male-typed domains, this would be insufficient when employers have a strong bias to favor men in the hiring process of male-typed domains (see, e.g., Reuben et al. 2014). Therefore, parallel interventions in the demand-side that make the hiring process

as much gender blind as possible are certainly desirable (see, e.g., Goldin and Rouse 2000).³⁷

Finally, our analysis and discussion sections were directed to understand how to promote women to compete more in a mixed-sex context. However, there are circumstances in which being competitive might not be desirable like, for example, participating in joint work in a mixed-sex team. From this perspective, our results reveal that men compete too much, not that women compete too little. Hence, for mixed-sex contexts in which cooperation is the most desirable attribute, the relevant question to be studied seems to be how can we persuade men to be less competitive rather than persuading women to be more competitive.

³⁷ Gender bias seems to be present not only in the hiring process but also in the promotion process. Leading organizations like McKinsey, BCG, Goldman Sachs, and Catalyst are increasingly searching for mechanisms to tackle this problem.

REFERENCES

- Almås, I., Cappelen, A., Salvanes, K., Sorensen, E., & Tungodden, B. (2016). Willingness to Compete: Family Matters. *Management Science*, 62(8), 2149-62.
- Balafoutas, L., & Sutter, M. (2012). Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory. *Science*, 335, 579-82.
- Bem, S. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88(4), 354-64.
- Blau, F., Ferber, M., & Winkler, A. (2013). *The Economics of Women, Men, and Work*. Englewood Cliffs, NJ: Prentice Hall.
- Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2), 542-55.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2014). Stereotypes. Working Paper.
- Brandts, J., Groenert, V., & Rott, C. (2014). The Impact of Advice on Women's and Men's Selection into Competition. *Management Science*, forthcoming.
- Buser, T., Dreber, A., & Mollerstrom, J. (2016). The impact of stress on tournament entry. *Experimental Economics*, forthcoming.
- Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, Competitiveness and Career Choices. *The Quarterly Journal of Economics*, 129(3), 1409-47.
- Cason, T., Masters, W., & Sheremeta, R. (2010). Entry into Winner-Take-All and Proportional-Prize Contests: An Experimental Study. *Journal of Public Economics*, 94(9-10), 604-11.
- Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2), 1-27.
- Dohmen, T., & Falk, A. (2011). Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender. *American Economic Review*, 101(2), 556-90.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. (2011). Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences. *Journal of the European Economic Association*, 9(3), 522-50.
- Fischbacher, U. (2007). z-Tree Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10(2), 171-78.
- Gneezy, U., Leonard, K., & List, J. (2009). Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica*, 77(5), 1637-64.
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, 118(3), 1049-74.
- Goldin, C., & Rouse, C. (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review*, 90(4), 715-41.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-25.

- Gupta, N., Poulsen, A., & Villeval, M. (2013). Gender matching and competitiveness: experimental evidence. *Economic Inquiry*, 51(1), 816-35.
- Inzlicht, M., & Schmader, T. (2012). *Stereotype threat: Theory, process, and application*. New York: Oxford Univ. Press.
- Kahneman, D., & Tversky, A. (2000). *Choice, Values, and Frames*. New York: Cambridge University Press.
- Kamas, L., & Preston, A. (2010). Are Women Really Less Competitive than Men? Working Paper.
- McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27, 415-44.
- Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics*, 122(3), 1067-1101.
- Niederle, M., & Vesterlund, L. (2011). Gender and Competition. *Annual Review in Economics*, 3, 601-30.
- Niederle, M., Segal, C., & Vesterlund, L. (2013). How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness. *Management Science*, 59(1), 1-16.
- Reuben, E., Sapienza, P., & Zingales, L. (2014). How Stereotypes Impair Women's Career in Science. *Proceedings of the National Academy of Sciences*, 111(12), 4403-08.
- Reuben, E., Wiswall, M., & Zafar, B. (2013). Preferences and Biases in Educational Choices and Labor Market Expectations: Shrinking the Black Box of Gender. Working Paper.
- Spencer, S., Logel, C., & Davies, P. (2016). Stereotype threat. *Annual review of psychology*, 67, 415-37.
- Spencer, S., Steele, C., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4-28.
- Sutter, M., & Rützler, D. (2014). Gender Differences in the Willingness to Compete Emerge Early in Life and Persist. *Management Science*, forthcoming.
- Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Connecticut (USA): Yale University Press.
- Wozniak, D., Harbaugh, W., & Mayr, U. (2014). The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices. *Journal of Labor Economics*, 32(1), 161-98.
- Zizzo, D. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75-98.

Appendix 2.A.1. Additional data

2.A.1.1. Probit models of choice of sex

Table 2.A.1.1. Probit models of choice of sex

Note: The table reports marginal effects. *Gap in winning belief* equals: [(winning belief against women) - (winning belief against men)]. Robust standard errors in brackets.

Dependent variable: 1 if the choice is a female competitor		
	Men (1)	Women (2)
<i>Baseline performance</i>	-0.006 [0.016]	-0.003 [0.015]
<i>Risk score</i>	-0.03 [0.035]	0.03 [0.031]
<i>Gap in winning belief</i>	0.004 [0.006]	0.007 [0.007]
Observations	34	30
Pseudo R ²	0.026	0.069

2.A.1.2. Men's perception of the gender composition

Table 2.A.1.2. Men's perception of the gender composition (in percentage)

Note: In each condition, the percentage are computed relatively to the women who reveal to have noticed the gender composition.

	Only female	Mainly female	Balanced composition	Mainly male	Only male
Replication	0%	12%	69%	19%	0%
Choice of Sex	0%	5%	76%	19%	0%

2.A.1.3. Men's beliefs about the likelihood of each sex to enter competition

Table 2.A.1.3. Men's beliefs about the likelihood of each sex to enter competition

Note: **Belief probability men enter** equals: (estimate of male participants who enter competition/total male participants in the session-1)*100; **Belief probability women enter** equals: [estimate of female participants who enter competition/(total female participants in the session)]*100. We report the average of these variables.

	Replication	Choice of Sex
Belief probability men enter	68.8% (17.7)	71.3% (15.0)
Belief probability women enter	43.6% (22.1)	44.6% (18.7)

Appendix 2.A.2.

The choice of payment scheme—the normative view

2.A.2.1. Expected payoff maximization

Consider an arbitrary participant i of the experiment who maximizes expected payoff and let x_i represent the number of correct addition problems worth € 1 each. Participant i 's preferences over monetary outcomes are represented by a twice-differentiable utility function $u_i : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$. Assume that $u_i(0) = 0$ and $u_i'(x_i) > 0$ for $x_i \geq 0$. Moreover, participant i 's cost to solve addition problems correctly is represented by a strictly convex function $c_i : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$. Assume that $c_i(0) = 0$, $c_i'(0) = 0$, $c_i'(x_i) > 0$ for $x_i > 0$, $c_i'(\infty) = \infty$, and $c_i''(x_i) > 0$ for $x_i \geq 0$.

For the Round 2 performance, participant i has to choose between a non-competitive and a competitive payment scheme. If participant i chooses the non-competitive payment scheme for the Round 2 performance, participant i faces the following problem:

$$(1) \quad \max_{x_i} \pi_i^{nc}(x_i) \quad \text{with} \quad \pi_i^{nc}(x_i) = 0.5 \cdot u_i(x_i) - c_i(x_i)$$

in which 0.5 represents the probability of participant i being paid for the subsequent performance.

Alternatively, if participant i chooses the winner-take-all tournament payment scheme for the Round 2 performance, participant i faces the following problem:

$$(2) \quad \max_{x_i} \pi_i^{com}(x_i | \bar{x}_i^1, b_{i,j}^1) \quad \text{with} \quad \pi_i^{com}(x_i | \bar{x}_i^1, b_{i,j}^1) = bwin_i(\bar{x}_i^1, b_{i,j}^1) \cdot u_i(x_i) - c_i(x_i)$$

in which $bwin_i(\bar{x}_i^1, b_{i,j}^1)$ represents participant i 's belief in the probability of winning the competition³⁸, \bar{x}_i^1 represents participant i 's number of correct

³⁸ As we explain in section 2.4.2., we consider $bwin_i(\bar{x}_i^1, b_{i,j}^1)$ rather than $bwin_i(x_i, b_{i,j})$ —in which x_i represents participant i 's number of correct answers in the subsequent Round 2 performance, and $b_{i,j}$ represents participant i 's belief about the probability distribution of the competitor j 's number of correct

answers in the Round 1 baseline performance, and $b_{i,j}^1$ represents participant i 's belief about the probability distribution of the competitor j 's number of correct answers in the Round 1 baseline performance.

In the replication and all women conditions, $bwin_i(\bar{x}_i^1, b_{i,j}^1)$ represents the belief in the probability of winning the competition against a participant randomly chosen among the participants who have chosen to compete. In the choice of sex condition, participant i considers two versions of problem (2). In one version, $bwin_i(\bar{x}_i^1, b_{i,j}^1)$ represents the belief in the probability of winning the competition against a man randomly chosen among the men who have chosen to compete. In the other version, $bwin_i(\bar{x}_i^1, b_{i,j}^1)$ represents the belief in the probability of winning the competition against a woman randomly chosen among the women who have chosen to compete. In this condition, participant i 's winning belief corresponds to the highest $bwin_i(\bar{x}_i^1, b_{i,j}^1)$ of the two versions of problem (2).

Proposition 1. *Suppose $u_i(x_i)$ is linear or strictly concave. Then,*

- (a) $\max_{x_i} \pi_i^{com}(x_i | \bar{x}_i^1, b_{i,j}^1)$ has a unique solution, denoted by x_i^{com*} , which is positive.
- (b) $\max_{x_i} \pi_i^{nc}(x_i)$ has a unique solution, denoted by x_i^{nc*} , which is positive.

Proposition 2. *Suppose $u_i(x_i)$ is linear or strictly concave. Then:*

- (a) If $bwin_i(\bar{x}_i^1, b_{i,j}^1) > 0.5$, then $\pi_i^{com}(x_i^{com*}) > \pi_i^{nc}(x_i^{nc*})$.
- (b) If $bwin_i(\bar{x}_i^1, b_{i,j}^1) < 0.5$, then $\pi_i^{com}(x_i^{com*}) < \pi_i^{nc}(x_i^{nc*})$.
- (c) If $bwin_i(\bar{x}_i^1, b_{i,j}^1) = 0.5$, then $\pi_i^{com}(x_i^{com*}) = \pi_i^{nc}(x_i^{nc*})$.

In short, for a risk neutral or risk averse participant, we only need to learn a participant's winning belief to determine whether a participant's choice of payment scheme is consistent with expected payoff maximization.

2.A.2.2 Proof of the propositions

Proof of Proposition 1

Proof of claim (a). In case participant i chooses the competitive payment scheme, participant i 's problem is:

answers in the subsequent Round 2 performance— because eliciting $bwin_i(x_i, b_{i,j})$ poses a serious incentive problem.

$$\max_{x_i} \pi_i^{com}(x_i | \bar{x}_i^1, b_{i,j}^1) \text{ with } \pi_i^{com}(x_i | \bar{x}_i^1, b_{i,j}^1) = bwin_i(\bar{x}_i^1, b_{i,j}^1) \cdot u_i(x_i) - c_i(x_i)$$

The f.o.c. is:

$$\pi_i'^{com}(x_i | \bar{x}_i^1, b_{i,j}^1) = 0 \Leftrightarrow bwin_i(\bar{x}_i^1, b_{i,j}^1) \cdot u_i'(x_i) = c_i'(x_i)$$

For $x_i = 0$, we have that $bwin_i(\bar{x}_i^1, b_{i,j}^1) \cdot u_i'(0) > c_i'(0)$ because $bwin_i(\bar{x}_i^1, b_{i,j}^1) > 0$, $u_i'(0) > 0$, and $c_i'(0) = 0$. Hence, since $u_i''(x_i) \leq 0$ and $c_i''(x_i) > 0$ for $x_i > 0$, this implies that there exists a unique $x_i > 0$ such that $bwin_i(\bar{x}_i^1, b_{i,j}^1) \cdot u_i'(x_i) = c_i'(x_i)$ because $u_i'(x_i)$ is non-increasing whereas $c_i'(x_i)$ is strictly increasing for $x_i > 0$ and $c_i'(\infty) = \infty$. We conclude, therefore, that $\pi_i^{com}(x_i | \bar{x}_i^1, b_{i,j}^1)$ has a unique stationary point, which we denote by x_i^s .

To verify that x_i^s is the argument of a maximum, we evaluate the s.o.c. for $x_i = x_i^s$:

$$\pi_i''^{com}(x_i | \bar{x}_i^1, b_{i,j}^1) \Big|_{x_i = x_i^s} = bwin_i(\bar{x}_i^1, b_{i,j}^1) \cdot u_i''(x_i^s) - c_i''(x_i^s) < 0$$

Because $bwin_i(\bar{x}_i^1, b_{i,j}^1) > 0$, $u_i''(x_i^s) \leq 0$, and $c_i''(x_i^s) > 0$.

Hence, $x_i^s = x_i^{com*}$ is indeed the argument of a maximum. QED

Proof of claim (b). The proof of claim (b) is analogous to the proof of claim (a).

Proof of Proposition 2

We begin by noting that $u_i(x_i^{nc*}) > 0$ and $u_i(x_i^{com*}) > 0$ because $u_i(0) = 0$ and $u_i'(x_i) > 0$ for $x_i > 0$.

Proof of claim (a). Consider $\pi_i^{com}(x_i^{nc*})$ and $\pi_i^{nc}(x_i^{nc*})$, which represent participant i 's expected payoff under the competitive and non-competitive payment scheme, respectively, in case participant i correctly solves x_i^{nc*} addition problems. Then:

$$\begin{aligned}
 \pi_i^{com}(x_i^{nc*}) - \pi_i^{nc}(x_i^{nc*}) &= bwin_i(\bar{x}_i^1, b_{i,j}^1) \cdot u_i(x_i^{nc*}) - c_i(x_i^{nc*}) - \left[0.5 \cdot u_i(x_i^{nc*}) - c_i(x_i^{nc*}) \right] \\
 &= u_i(x_i^{nc*}) \cdot \left[bwin_i(\bar{x}_i^1, b_{i,j}^1) - 0.5 \right]
 \end{aligned}$$

Hence, if $bwin_i(\bar{x}_i^1, b_{i,j}^1) > 0.5$ we have that $u_i(x_i^{nc*}) \cdot \left[bwin_i(\bar{x}_i^1, b_{i,j}^1) - 0.5 \right] > 0$ because $u_i(x_i^{nc*}) > 0$. This implies that:

$$(3) \quad \pi_i^{com}(x_i^{nc*}) > \pi_i^{nc}(x_i^{nc*})$$

Moreover, as x_i^{com*} is the argument max of $\pi_i^{com}(x_i)$, we necessarily have that $\pi_i^{com}(x_i^{com*}) \geq \pi_i^{com}(x_i^{nc*})$ ³⁹, which combined with (3) implies that $\pi_i^{com}(x_i^{com*}) > \pi_i^{nc}(x_i^{nc*})$ when $bwin_i(\bar{x}_i^1, b_{i,j}^1) > 0.5$.

Proof of claim (b). The proof of claim (b) is analogous to the proof of claim (a).

Proof of claim (c). If $bwin_i(\bar{x}_i^1, b_{i,j}^1) = 0.5$ participant i faces the same problem under either the competitive or the non-competitive payment scheme, which trivially means that $x_i^{com*} = x_i^{nc*}$. This necessarily implies that $\pi_i^{com}(x_i^{com*}) = \pi_i^{nc}(x_i^{nc*})$ when $bwin_i(\bar{x}_i^1, b_{i,j}^1) = 0.5$. QED

³⁹ In fact, $\pi_i^{com}(x_i^{com*}) > \pi_i^{com}(x_i^{nc*})$ because $x_i^{com*} \neq x_i^{nc*}$ if $bwin_i(\bar{x}_i^1, b_{i,j}^1) > 0.5$.

Appendix 2.A.3.

Additional notes on the experimental procedure

2.A.3.1. Equal number of male and female participants in the replication and choice of sex conditions

In the replication and choice of sex conditions, our objective was to conduct perfectly gender balanced sessions, up to a max of 15 men and 15 women per session.

The procedure we used to ensure an equal number of male and female participants per session was:

1. We separately invite men and women via ORSEE. We allow a max of 20 men and 20 women to register per session, respectively.
2. When a prospective participant is arriving for a session, they find experimenter 1, who is standing just outside the lab's entrance door to check a prospective participant's ID. Experimenter 1 ensures in a subtle way that a max of 15 men and 15 women enter the lab's welcome room.
3. Once entering the welcome room, a prospective participant immediately finds experimenter 2, who is standing right behind of the entrance's door. Experimenter 2 provides a prospective participant with a numbered card while inviting them to take a seat in the welcome room. Experimenter 2 ensures in a subtle way that women (men) receive numbered cards ranging from 1 to W (16 to 15+M), in which $W \leq 15$ and $M \leq 15$ represent the number of women and men who enter the welcome room, respectively.
4. Inside of the welcome room, there is also experimenter 3 who just observes the arrival of prospective participants until receiving a sign from experimenter 2. The sign means that the door has been locked because either 30 prospective participants are already in the room or, although there are less than 30 prospective participants in the room, the planned time to allow the entrance of prospective participants into the welcome room has elapsed.
5. After quickly checking the number of prospective participants in the room, experimenter 3:

- a. Starts right away the welcome speech in case there is an equal number of male and female participants in the room (15-15, 14-14, and so on).
- b. Otherwise, announces: “Due to scientific purposes, we need a precise number of participants to run this experimental session, and unfortunately you are too many. Thus, some of you will not be able to participate in the experiment. Nevertheless, who will be invited to leave will be paid € 3 for showing up on time today”. The procedure to determine who leaves the room is performed in a way to look random within all prospective participants, although in reality we only randomize within the prospective participants whose sex is overrepresented. Specifically, consider that $W - M = \kappa > 0$. Then, experimenter 3 randomly draws a number from a pot that contains W numbers ranging from 1 to W and repeats the procedure $\kappa - 1$ more time(s) without replacement. Once it is determined the person(s) that need to leave, experimenter 3 politely asks the person(s) to meet outside experimenter 1 to receive the show-up fee. Only after this person(s) leave the room, experimenter 3 starts the welcome speech. An analogous procedure is performed if instead $M > W$.

Outcome: All sessions were filled up with 15 men and 15 women, except one session of the replication condition that was conducted with 14 men and 14 women.

2.A.3.2. Forming the competition pairs in the choice of sex condition

Participants who choose competition are matched to participants with equivalent choices. This means that three types of pairs are possible:

1. ***Single-sex competition pair of men:*** a man who chooses a male competitor is matched to another man who also chooses a male competitor.

In case an odd number of men choose to compete against a man, a randomly chosen performance among the men who have been matched in single-sex pairs is used a second time to determine the earnings of the unmatched man. If only one man chooses to compete against a man (something that did not happen), the competition is not implemented, and the male participant is paid according to the non-competitive payment.

2. ***Single-sex competition pair of women:*** a woman who chooses a female competitor is matched to another woman who also chooses a female competitor.

In case an odd number of women choose to compete against a woman, a randomly chosen performance among the women who have been matched in single-sex pairs is used a second time to determine the earnings of the unmatched woman. If only one woman chooses to compete against a woman

(something that did not happen), the competition is not implemented, and the female participant is paid according to the non-competitive payment.

3. ***Mixed-sex competition pair.*** A man (woman) who chooses a female (male) competitor is matched to a woman (man) who chooses a male (female) competitor.

In case $M(w) - W(m) = a > 0$ and $W(m) > 0$ (in which $M(w)$ represents the number of men who choose a female competitor and $W(m)$ represents the number of women who choose a male competitor), a randomly chosen performance among $W(m)$ is used a more time(s) to determine the earnings of the unmatched man (men). If $M(w) > W(m)$ but $W(m) = 0$ (something that did not happen), the competition is not implemented and the male participant(s) is/are paid according to the non-competitive payment. An analogous procedure is performed if instead $W(m) > M(w)$ (something that did not happen).

Appendix 2.A.4.

Experimental instructions

Welcome speech [Given by one experimenter in the welcome room before participants move into the computers' room]

Good morning/afternoon! I welcome you to this experiment. My name is Diogo Geraldes, and I am the organizer of today's experiment. This experiment is part of a research project on decision-making financed by a scientific institution. The approximate duration of the experiment is 1 hour and 30 minutes. All the information provided in this experiment is truthful. If you wish, you can inspect this at the end of the experiment.

In the lab, you can earn money with the decisions you make. The amount you earn will be paid out in cash, individually and confidentially, at the end of the experiment.

On the card you are holding, you see a number. In the lab, the computer cubicles are also numbered. You will take a seat in the computer cubicle with the same number as your card.

I ask you now a moment before you enter the computer room (quick check of z-leaf units and number of participants)

You are about to enter the computer room and take a seat in the computer cubicle that corresponds to your number. We use these numbers for the data analysis. The number will not be linked to your name. Your anonymity is therefore secured.

In the lab, you will receive detailed instructions about the experiment. It is for us important that you understand and follow the instructions carefully. In particular, I ask you not to communicate in any other way than specified in the instructions.

Before you enter the lab, please switch off your mobile phone and store it inside your bag/pocket.

Are there any questions at this moment? Please, you can enter now into the computer room.

Welcome screen [*same version in the three conditions*]

Welcome to the BEElab!

You will now take part in a decision-making experiment.

Please click on <continue>.

<Continue>

General instructions [*same version in the three conditions*]

You are now taking part in a decision-making experiment. For showing up on time today you will be paid € 3. In addition, you can earn money with the decisions you make. Hence, it is important that you fully understand the instructions that follow. Please read them carefully.

In this experiment, you will experience two rounds (ROUND 1 and ROUND 2), each with a work task. Each work task will take 5 minutes. Before each round, we will describe in detail how your earnings are determined. At the end of the experiment, we will randomly select only one of the rounds for payment. You will be paid according to your earnings in that round.

Moreover, you will be asked some questions during the experiment. You can earn additional money with the answers to some of these questions.

Your total earnings from the experiment are, therefore, the sum of your payment for the randomly selected round, the sum of your payment for the questionnaires, and the € 3 show up fee. You will be paid out in cash at the end of the experiment. The payment will take place individually and confidentially.

Before you perform ROUND 1, there is an unpaid PRACTICE ROUND to make you familiar with the work task with which you can earn money in the subsequent ROUND 1 and ROUND 2.

If you understand these instructions, please click on the OK-button to read the PRACTICE ROUND instructions. Otherwise, please raise your hand. An experimenter will then come to you and answer your questions.

<OK>

Practice round [*same version in the three conditions*]

This PRACTICE ROUND serves to familiarize you with the work task with which you can earn money in the subsequent ROUND 1 and ROUND 2. In this PRACTICE ROUND, you are not paid. However, it is in your best interest to take this round seriously because the more practice you gain now, the more money you are likely to earn in the subsequent paid rounds.

In this round, you will be asked to calculate the sum of 5 two-digit numbers (e.g.: $84 + 94 + 17 + 19 + 39 = ?$). You will be given 2 minutes to calculate as many of these addition problems as possible.

Please use the keyboard to type in your answer into the box and click the "OK"-button to confirm it. The computer will immediately tell you whether your answer was correct or wrong and a new addition problem will instantaneously be displayed on your screen. Your answers to the problems are anonymous.

You must answer the addition problems without any helping device such as a pocket calculator or a mobile phone. You are welcome, though, to make use of the provided scratch paper in case you wish to write the numbers down.

The remaining time in seconds will be shown on the top right of the screen. If you have a question, please raise your hand.

Please click on the OK-button to start the PRACTICE ROUND.

<OK>

Round 1 - Piece Rate *[same version in the three conditions]*

In this round, you will be given 5 minutes to calculate as many addition problems as possible. The procedure to submit your answers is the same as before.

Recall that only one round (ROUND 1 or ROUND 2) is randomly selected for payment. If this round is selected, you get € 0.50 per problem that you correctly solve. Incorrectly solved problems do not count (neither positively nor negatively). We refer to this payment as the *piece rate* payment.

Example 1: If you correctly solve 35 addition problems in the 5 minutes and this round is the one randomly selected for payment, then you will earn: € 17.5 = 35 x € 0.50

Example 2: If you correctly solve 4 addition problems in the 5 minutes and this round is the one randomly selected for payment, then you will earn: € 2 = 4 x € 0.50

The numbers in the examples are just for illustrative purposes. They DO NOT intend to suggest how anyone should perform in this experiment.

The remaining time in seconds will be shown on the top right of the screen. If you have a question, please raise your hand.

Please click on the OK-button to start the ROUND 1.

<OK>

Round 2 – Choice of payment scheme for a subsequent performance *[version of the replication and all women conditions]*

Screen 1

As in ROUND 1, you will be given 5 minutes to calculate as many addition problems as possible. The procedure to submit your answers is the same as before.

However, you will now be able to choose the payment scheme you prefer to apply to your performance in ROUND 2. In particular, you can choose between two alternative payment schemes.

If ROUND 2 is randomly selected for payment, then your earnings for this round are determined as follows.

1. NON-COMPETITIVE PAY OPTION: If you choose this option you will be randomly paired with another participant who has also chosen the NON-COMPETITIVE PAY. At the end of the experiment, one of you two will be chosen randomly with a 50% probability. If you are the chosen one, you will earn € 1 for each correctly solved addition problem. If your paired participant is the chosen one, you will earn nothing.

Example 1: If you correctly solve 35 addition problems in ROUND 2 and you are the one randomly chosen at the end of the experiment, you will earn € 35 = 35 x € 1.

Example 2: If the participant with whom you have been paired is the one randomly chosen at the end of the experiment, you will earn 0 whatever your number of correctly solved addition problems in ROUND 2 is.

2. TOURNAMENT OPTION: If you choose this option you will compete against a participant randomly chosen among the participants who also have chosen to compete.

The competitive scheme works as follows: If you correctly solve more addition problems in ROUND 2 than your paired participant, you will earn € 1 for each addition problem you correctly solve. In case you and your paired participant correctly solve the same number of addition problems in ROUND 2 then one of you is chosen as the winner with a 50% probability. If you correctly solve less addition problems in ROUND 2 than your paired participant, you will earn nothing.

Example 1: If you correctly solve 4 addition problems in ROUND 2 and the participant with whom you have been randomly paired correctly solves less than 4 addition problems, you are the better performer, and you will earn € 4 = 4 x € 1.

Example 2: If you correctly solve 4 addition problems in ROUND 2 and the participant with whom you have been randomly paired correctly solves more than 4 addition problems, you are the worse performer, and you will earn nothing.

Example 3: If both of you correctly solve 4 addition problems in ROUND 2, then you will earn € 4 with a 50% probability.

The numbers in the examples are just for illustrative purposes. They DO NOT intend to suggest how anyone should perform in this experiment.

<Continue ROUND 2 instructions>

Screen 2

You will not be informed about the identity of the participant you have been paired with, neither during nor after the experiment. You will also not be informed about how many addition problems this participant correctly solved in

this round. The same applies to your paired participant. Your choice of payment scheme will stay anonymous.

NOTES:

- i. If you choose to compete, but the other participants' choices in this experiment are such that you cannot be paired with anyone, the tournament will not take place. If this unlikely situation arises, you will be informed after you made your choice and you will be paid in this round according to the NON-COMPETITIVE PAY option.
- ii. If you are the only participant in this experiment who chooses the NON-COMPETITIVE PAY option, you will still be paid according to this option. That is, at the end of the experiment you will be chosen with a 50% probability to be paid. If you get chosen you will earn € 1 for each addition problem, you correctly solved in ROUND 2.
- iii. The new sequence of addition problems in ROUND 2 is the same for all participants. That is, all participants in this experiment will see the same sequence of addition problems regardless of their choice of payment scheme.

The next computer screen will ask you to choose whether you want the non-competitive pay or the tournament applied to your subsequent ROUND 2 performance. If you have a question, please raise your hand.

Please click on the OK-button to make your choice.

<Back to the previous screen> <OK>

Round 2 – Choice of payment scheme for a subsequent performance
[version of the choice of sex condition]

As in ROUND 1, you will be given 5 minutes to calculate as many addition problems as possible. The procedure to submit your answers is the same as before.

However, you will now be able to choose the payment scheme you prefer to apply to your performance in ROUND 2. In particular, you can choose between three alternative payment schemes.

If ROUND 2 is randomly selected for payment, then your earnings for this round are determined as follows.

1. NON-COMPETITIVE PAY OPTION: If you choose this option you will be randomly paired with another participant who has also chosen the NON-COMPETITIVE PAY. At the end of the experiment, one of you two will be chosen randomly with a 50% probability. If you are the chosen one, you will earn € 1 for each correctly solved addition problem. If your paired participant is the chosen one, you will earn nothing.

Example 1: If you correctly solve 35 addition problems in ROUND 2 and you are the one randomly chosen at the end of the experiment, you will earn € 35 = 35 x € 1.

Example 2: If the participant with whom you have been paired is the one randomly chosen at the end of the experiment, you will earn 0 whatever your number of correctly solved addition problems in ROUND 2 is.

2. TOURNAMENT AGAINST A MAN OPTION: If you choose this option you will compete against a male participant randomly chosen among the male participants who also have chosen to compete.

3. TOURNAMENT AGAINST A WOMAN OPTION: If you choose this option you will compete against a female participant randomly chosen among the female participants who also have chosen to compete.

The only difference between options 2 and 3 is whether your paired opponent is a male or a female participant. The competitive scheme is the same in both options, and works as follows: If you correctly solve more addition problems in ROUND 2 than your paired participant, you will earn € 1 for each addition problem you correctly solve. In case you and your paired participant correctly solve the same number of addition problems in ROUND 2 then one of you is chosen as the winner with a 50% probability. If you correctly solve less addition problems in ROUND 2 than your paired participant, you will earn nothing.

Example 1: If you correctly solve 4 addition problems in ROUND 2 and the participant with whom you have been randomly paired correctly solves less than 4 addition problems, you are the better performer, and you will earn € 4 = 4 x € 1.

Example 2: If you correctly solve 4 addition problems in ROUND 2 and the participant with whom you have been randomly paired correctly solves more than 4 addition problems, you are the worse performer, and you will earn nothing.

Example 3: If both of you correctly solve 4 addition problems in ROUND 2, then you will earn € 4 with a 50% probability.

The numbers in the examples are just for illustrative purposes. They DO NOT intend to suggest how anyone should perform in this experiment.

<Continue ROUND 2 instructions>

Screen 2

You will not be informed about the identity of the participant you have been paired with, neither during nor after the experiment. You will also not be informed about how many addition problems this participant correctly solved in this round. The same applies to your paired participant. Your choice of payment scheme will stay anonymous.

NOTES:

i. If you choose to compete, but the other participants' choices in this experiment are such that you cannot be paired with anyone, the tournament will not take

place. If this unlikely situation arises, you will be informed after you made your choice and you will be paid in this round according to the NON-COMPETITIVE PAY option.

ii. If you are the only participant in this experiment who chooses the NON-COMPETITIVE PAY option, you will still be paid according to this option. That is, at the end of the experiment you will be chosen with a 50% probability to be paid. If you get chosen you will earn € 1 for each addition problem, you correctly solved in ROUND 2.

iii. The new sequence of addition problems in ROUND 2 is the same for all participants. That is, all participants in this experiment will see the same sequence of addition problems regardless of their choice of payment scheme.

The next computer screen will ask you to choose whether you want a non-competitive pay, a tournament against a man or a tournament against a woman applied to your subsequent ROUND 2 performance. If you have a question, please raise your hand.

Please click on the OK-button to make your choice.

<Back to the previous screen> <OK>

Winning beliefs [*version of the replication and all women conditions*]

Before the start of ROUND 2, please provide your best estimate on the following question using an integer number between 0 and 100 (inclusive). You can earn additional money with your answer.

Consider ONLY the OTHER participants who have just chosen to compete in the subsequent performance round. How many of these participants (in percentage terms) did correctly solve LESS addition problems than you in the 5 minutes ROUND 1 piece-rate performance?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual percentage (rounded to the closest integer number). If your estimate deviates by at most plus or minus 1 to the actual percentage, you will earn € 4. If your estimate deviates by more than plus or minus 1 but at most plus or minus 4 from the actual percentage, you will earn € 2. If your estimate deviates by more than plus or minus 4 but at most plus or minus 7 from the actual percentage, you will earn € 1. If your estimate deviates by more than plus or minus 7 from the actual percentage, you will earn nothing.

Please fill in your best estimate here (integer number):

<OK>

Winning beliefs [*version of the choice of sex condition*]

Before the start of ROUND 2, please provide your best estimate on the following two questions using an integer number between 0 and 100 (inclusive). You can earn additional money with your answers.

Consider ONLY the OTHER MALE participants who have just chosen to compete in the subsequent performance round. How many of these male participants (in percentage terms) did correctly solve LESS addition problems than you in the 5 minutes ROUND 1 piece-rate performance?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual percentage (rounded to the closest integer number). If your estimate deviates by at most plus or minus 1 to the actual percentage, you will earn € 4. If your estimate deviates by more than plus or minus 1 but at most plus or minus 4 from the actual percentage, you will earn € 2. If your estimate deviates by more than plus or minus 4 but at most plus or minus 7 from the actual percentage, you will earn € 1. If your estimate deviates by more than plus or minus 7 from the actual percentage, you will earn nothing.

Please fill in your best estimate for male participants here (integer number):

Consider ONLY the OTHER FEMALE participants who have just chosen to compete in the subsequent performance round. How many of these female participants (in percentage terms) did correctly solve LESS addition problems than you in the 5 minutes ROUND 1 piece-rate performance?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual percentage (rounded to the closest integer number). If your estimate deviates by at most plus or minus 1 to the actual percentage, you will earn € 4. If your estimate deviates by more than plus or minus 1 but at most plus or minus 4 from the actual percentage, you will earn € 2. If your estimate deviates by more than plus or minus 4 but at most plus or minus 7 from the actual percentage, you will earn € 1. If your estimate deviates by more than plus or minus 7 from the actual percentage, you will earn nothing.

Please fill in your best estimate for female participants here (integer number):

<OK>

Perception of the gender composition [same version in the three conditions]

How do you perceive the gender composition of participants in this experimental session?

Please choose the option that better describes your perception:

- ☐ I did not notice the gender composition of participants
- ☐ Only female participants

- ☐ Mainly female participants
- ☐ Balanced composition
- ☐ Mainly male participants
- ☐ Only male participants

<OK>

Beliefs about the inclination of each sex to enter competition [*version of the replication and choice of sex conditions*]

Please provide your best estimate on the following two questions using an integer number between 0 and 15 (inclusive). You can earn additional money with your answers.

For your information, there are 15 male participants in this experimental session. How many male participants do you think have chosen to compete (excluding yourself if you are a male participant and chose competition)?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual number of male participants who have chosen to compete. If your estimate is equal to the actual number, you will earn € 2. If your estimate deviates by plus or minus 1 from the actual number, you will earn € 1. If your estimate deviates by plus or minus 2 from the actual number, you will earn € 0.5. If your estimate deviates by more than plus or minus 2 from the actual number, you will earn nothing.

Please fill in your best estimate for male participants here (integer number):

For your information, there are 15 female participants in this experimental session. How many female participants do you think have chosen to compete (excluding yourself if you are a female participant and chose competition)?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual number of female participants who have chosen to compete. If your estimate is equal to the actual number, you will earn € 2. If your estimate deviates by plus or minus 1 from the actual number, you will earn € 1. If your estimate deviates by plus or minus 2 from the actual number, you will earn € 0.5. If your estimate deviates by more than plus or minus 2 from the actual number, you will earn nothing.

Please fill in your best estimate for female participants here (integer number):

<OK>

Belief about the inclination of participants to enter competition [*version of the all women condition*]

Please provide your best estimate on the following question using an integer number between 0 and 30 (inclusive). You can earn additional money with your answers.

For your information, there are 30 female participants in this experimental session. How many female participants do you think have chosen to compete (excluding yourself if you chose competition)?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual number of female participants who have chosen to compete. If your estimate is equal to the actual number, you will earn € 4. If your estimate deviates by plus or minus 1 from the actual number, you will earn € 2. If your estimate deviates by plus or minus 2 from the actual number, you will earn € 1. If your estimate deviates by more than plus or minus 2 from the actual number, you will earn nothing.

Please fill in your best estimate here (integer number):

Stereotype-based beliefs *[version of the replication and choice of sex conditions]*

FOR YOUR INFORMATION: In previous studies that used the same five two-digit numbers addition task, it was found that the minimum number of addition problems correctly solved during 5 minutes by any participant was 4 and the maximum number was 35. If you wish, you can inspect these studies after the experiment

Please provide your best estimate on the following two questions using an integer number between 4 and 35 (inclusive). You can earn additional money with your answers.

How many addition problems did male participants correctly solve, on average, in the 5 minutes ROUND 1 piece rate performance (excluding yourself if you are a man)?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual average number of addition problems (rounded to the closest integer number). If your estimate is equal to the actual average number, you will earn € 4. If your estimate deviates by plus or minus 1 from the actual average number, you will earn € 2. If your estimate deviates by plus or minus 2 from the actual average number, you will earn € 1. If your estimate deviates by more than plus or minus 2 from the actual average number, you will earn nothing.

Please fill in your best estimate for male participants here (integer number):

How many addition problems did female participants correctly solve, on average, in the 5 minutes ROUND 1 piece rate performance (excluding yourself if you are a woman)?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual average number of addition problems (rounded to the closest integer number). If your estimate is equal to the actual average number, you will earn € 4. If your estimate deviates by plus or minus 1 from the actual average number, you will earn € 2. If your estimate deviates by plus or minus 2 from the actual average number, you will earn € 1. If your estimate deviates by more than plus or minus 2 from the actual average number, you will earn nothing.

Please fill in your best estimate for female participants here (integer number):

<OK>

Stereotype-based beliefs [*version of the all women condition*]

FOR YOUR INFORMATION: In previous studies that used the same five two-digit numbers addition task, it was found that the minimum number of addition problems correctly solved during 5 minutes by any participant was 4 and the maximum number was 35. If you wish, you can inspect these studies after the experiment

Please provide your best estimate on the following two questions using an integer number between 4 and 35 (inclusive). You can earn additional money with your answers.

Reference to a previous experimental session in which there were male and female participants: How many addition problems did the male participants correctly solve, on average, in the 5 minutes ROUND 1 piece rate performance (excluding yourself if you are a man)?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual average number of addition problems (rounded to the closest integer number). If your estimate is equal to the actual average number, you will earn € 4. If your estimate deviates by plus or minus 1 from the actual average number, you will earn € 2. If your estimate deviates by plus or minus 2 from the actual average number, you will earn € 1. If your estimate deviates by more than plus or minus 2 from the actual average number, you will earn nothing.

Please fill in your best estimate for male participants here (integer number):

Reference to a previous experimental session in which there were male and female participants: How many addition problems did the female participants correctly

solve, on average, in the 5 minutes ROUND 1 piece rate performance (excluding yourself if you are a woman)?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual average number of addition problems (rounded to the closest integer number). If your estimate is equal to the actual average number, you will earn € 4. If your estimate deviates by plus or minus 1 from the actual average number, you will earn € 2. If your estimate deviates by plus or minus 2 from the actual average number, you will earn € 1. If your estimate deviates by more than plus or minus 2 from the actual average number, you will earn nothing.

Please fill in your best estimate for female participants here (integer number):

<OK>

Risk attitude [*same version in the three conditions*]

How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

Please tick a box on the scale, in which the value 0 means 'not at all willing to take risks' and the value 10 means very willing to take risks':

Not at all willing to take
risks



Very willing to take
risks

<Continue>

Sex and performance under competition: Is there a stereotype threat shadow? An intriguing behavior of men¹

3.1. Introduction

A politically, socially, and economically important stylized fact about gender differences is the gap in wages and positions at the workplace. In 2010 women earned, on average, 17% less than men in the 28 European Union countries (EU-28). With reference to positions, in academics for instance, only 20% of the heads of institutions in the EU-28 with the capacity to deliver a Ph.D. were women in 2014.² If we look at CEO positions among top firms, the figures for women are even worse.³

One common explanation for these gender differences at the workplace is discrimination (e.g., Black and Strahan 2001). Another one is women's higher sensitivity to work-family conflicts and women's weaker negotiation behavior (e.g., Babcock and Laschever 2003). An alternative explanation has been suggested by recent studies in experimental economics: *men are more inclined to compete than women*. These studies show that even under tightly controlled and relatively abstract situations where men and women compete with each other, a difference

¹ This chapter is based on Geraldes, Riedl, and Strobel (2016), and is written using the *Singular they* as a gender-neutral singular pronoun.

² Source: European Commission (2016).

³ For an overview of gender differences in the labor market, see Blau, Ferber, and Winkler (2013).

in attitude towards competition between the sexes exists. This growing body of research addresses two main questions:

- I. Differences in preference for competitive environments, i.e., do men and women differently self-select into a competitive environment?
- II. Differences in performance under exogenously set competition, i.e., do men and women differently react in terms of performance when they have to compete?

Concerning self-selection into competitive environments a standard finding is that women shy away from competition when given the option to compete whereas men do not (e.g., Niederle and Vesterlund 2007; Gneezy, Leonard and List 2009; Cason, Masters, and Sheremeta 2010; Dohmen and Falk 2011; Booth and Nolen 2012; Gupta, Poulsen, and Villeval 2013). Regarding performance under competitive pressure, the sparse experimental economics literature shows that in mixed-sex groups men generally improve their performance under competition whereas this is not (or much less) the case for women (see Gneezy, Niederle, and Rustichini 2003 and Gneezy and Rustichini 2004). This chapter extends the latter line of research by examining the connection between men's and women's competitive performance and stereotype threats.

The first objective of this chapter is to experimentally investigate the extent to which stereotype threats are related to men's and women's performance under competitive pressure when they *have* to compete with each other. To this end, we conduct a controlled laboratory experiment to examine men's and women's performance of a mathematical task when they *have* to compete with each other not only under an implicit stereotype threat but also under an explicit stereotype threat. Moreover, we evaluate the effect not only of explicitly induced stereotype-based expectations against women but also of explicitly induced stereotype-based expectations against men.

We hypothesize that: i. The competition induces participants to exert higher effort. Thus, relative to a control non-competitive performance, we expect participants to improve performance under competition when they are not the target of a stereotype threat; ii. Stereotype threat counteracts the positive effect of competition on performance. Thus, the competitive performance of participants targeted by stereotype threat also depends on how they perceive the stereotype threat.

Our results do not corroborate the well-documented finding that women's performance is impaired in situations entailing a mathematical stereotype threat against women (Inzlicht and Schmader 2012). When we induce, either implicitly or explicitly, a mathematical stereotype threat against women, we observe that women improve performance under competition as much as men do. However, perhaps surprisingly, we find evidence for the adverse effect of stereotype threat when we induce a mathematical stereotype threat against men.

In this chapter, we also investigate the effort participants should strategically exert in our competitive setting, which consists of a one-to-one mixed-sex winner-take-all tournament. We find that for (common knowledge) beliefs such that participants do not believe in a considerable gap in ability between men and women, both men and women should exert effort during the competition. However, for (common knowledge) beliefs such that participants believe in a considerable gap in ability favoring men (women), only men (women) should exert effort during the competition. Notably, in the latter scenario, the equilibrium effort of the more able participant is higher than the effort level the participant believes is sufficient to win the competition.

To our knowledge, the first study to establish a possible connection between stereotype threat and exogenously set economic competition between the sexes is Günther, Ekinci, Schwier, and Strobel (2010). These authors argue that the finding that men perform better than women under competition is possibly flawed because previous studies employ real-effort tasks for which stereotype-based expectations about male superior performance do exist. They hypothesize that in mixed-sex groups women compete less in tasks for which men are stereotypically believed to perform better than women. This effect, though, should not be observed in gender-neutral tasks and stereotypically believed female tasks. Their results indicate that women indeed react less to competitive incentives for a male task, but also indicate that women react as strongly as men and more strongly than men in response to competitive pressure for a gender-neutral task and a female task, respectively.

The main argument of Günther et al. (2010) is that the way a task is described suffices to activate a stereotype threat. Therefore, their study implies that implicit stereotype threats can explain gender differences in performance under competition. A stereotype threat experience could be either implicitly or explicitly activated. *Implicit activation* of stereotype threat refers to situations where simply placing the targeted individual in a context where the associated stereotype is well known, but not explicitly highlighted, is sufficient to induce the targeted individual to experience the threat. *Explicit activation* of stereotype threat refers to situations where the targeted individual experiences the threat by explicitly confronting the individual with the stereotype (Smith and White 2002).

In real-world situations of competition between the sexes as, for example, women competing for academic positions in math-intensive fields or aiming top paid corporate positions, women are often made explicitly aware that these positions are dominated by men and/or that people expect men to be more successful than women in those domains. Thus, although evaluating the effect of implicitly induced stereotype-based expectations gives an important first insight, we consider of utmost importance to also study competition between the sexes in situations where we explicitly induce stereotype-based expectations.

Our work particularly relates to the psychology literature on stereotype threat. Stereotype threat is essentially a situational predicament where a member of a group is apprehensive by the possibility of confirming a negative stereotype about their group (Steele and Aronson 1995). This literature reports that stereotype threat undermines task performance of various groups across multiple domains. The effect has been shown, for example, amongst African-Americans (see, e.g., Steele and Aronson 1995) and Latinos (see, e.g., Schmader and Johns 2003) when compared to Caucasians on tests labelled as indicators of intellectual ability, amongst women when compared to men during tests evaluating mathematical ability (see, e.g., Spencer, Steele, and Quinn 1999), and amongst Caucasian men in math-tests when informed about Asian-Americans' superior ability in mathematics (Aronson, Lustina, Good, Keough, Steele, and Brown 1999).

Essentially, stereotype threat theory predicts that the performance gap between two groups should be different depending on whether it is measured in a situation where one of the groups is subject to stereotype threat or in a situation where no stereotype threat exists. We should highlight, though, that the focus of the psychology research on stereotype threat is on measuring performance within non-competitive situations.⁴

The remainder of the chapter is organized as follows. In section 3.2, we describe the design of the lab experiment. In section 3.3, we state our hypotheses. In section 3.4, we present the results of the lab experiment. In section 3.5, we normatively analyze effort provision in our experimental setting. In section 3.6, we discuss our results. We conclude in section 3.7.

3.2. Design of the lab experiment

3.2.1. Methods

We design a between-subject lab experiment with three competitive conditions. Yet, the fundamental feature of our experimental design is the within-subject dimension of each condition. Specifically, each condition consists of two stages. In the first stage, participants perform individually a real-effort task under a non-competitive payment scheme. In the second stage, we randomly pair each participant with a participant of the opposite sex to subsequently perform a real-effort task under a competitive payment scheme. Our main measure is the change in performance between stages of participants. This within-subject measure allows us to analyze the performance of a real-effort task controlling for differences in unobserved participants' individual characteristics, especially participants' ability to perform the real-effort task.⁵

⁴ For a review of the Stereotype threat literature see Inzlicht and Schmader (2012); Spencer, Logel, and Davies (2016).

⁵ Gneezy and Rustichini (2004) already used this method in a field experiment measuring performance under competition. Iriberry and Rey-Biel (2017) also use a within-subject measure of performance.

3.2.2. The real-effort task

The real-effort task consists of multiplying one- and two-digit numbers and, following Dohmen and Falk (2011), we implement five different levels of difficulty.⁶ We present the problems on computer screens. Participants have to type each answer into a box and confirm it by clicking an “OK”-button with the mouse. Each time an answer is provided, we inform whether the answer is correct. If the answer is correct, a new problem instantaneously appears on the screen. If the answer is wrong, the same problem has to be tackled until the correct solution is provided. This procedure serves to prevent participants from guessing. We inform participants that we do not allow any aid (calculator, paper and pencil, etc.) to solve the multiplication problems.

We employ a mathematical task because of the well documented evidence that there is a stereotype that men are better at mathematics (e.g., Spencer et al. 1999).⁷ Moreover, as we document in the following subsection, there is reliable information concerning gender differences in mathematical ability that we can use to induce a converse stereotype threat without deceiving participants.

3.2.3. Detailed design of the lab experiment

In Figure 3.1, we show the sequence of events in the lab experiment. We start by informing participants that the experiment consists of two performance stages, and inform participants that they can earn money in each performance stage, and that the possible earnings in the first stage are independent of their own and others’ behavior in the second stage. We also inform participants that we do not provide feedback on performance at the end of each stage, neither in absolute terms nor relative to others.

Figure 3.1. Chart of the lab experiment

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Multiplication task practice	First stage performance	Confidence level elicitation	Second stage performance	Risk attitude elicitation	Competitive attitude elicitation

▪ ***Multiplication task practice.*** The experiment starts with an unpaid practice time in which we ask participants to calculate as many multiplication problems as possible within 2 minutes. This step serves to familiarize participants with the real-effort task. We inform participants that it is in their best interest to gain practice because they can earn money with the multiplication task in the performance stages.

⁶ Examples: 11×9 (level 1); 3×32 (level 2); 6×43 (level 3); 4×68 (level 4); 7×89 (level 5).

⁷ In an online questionnaire, we find evidence that men and women from our pool of subjects indeed hold this stereotype. We present the results of the online questionnaire in section 3.3.

▪ **First stage performance.** In this stage, we elicit each participant performance under non-competitive monetary incentives. The performance in this stage serves as the baseline for comparison with the performance in the subsequent competitive second stage. Before participants start performing the multiplication task, we inform them that they have been randomly paired with another participant present in the lab, without making any reference to sexes at this moment. We then ask participants to perform the multiplication task for 5 minutes under a *random pay* incentive scheme. Specifically, we randomly choose with equal probability one participant in each pair to be paid out. The instructions participants read explain the payment scheme as follows: “In this stage, you have been randomly paired with another participant. At the end of the experiment one of you two is randomly chosen with equal probability. The chosen one earns € 0.40 for each correctly solved multiplication problem. The other earns nothing.”

▪ **Confidence level elicitation.** Immediately after the non-competitive first stage, we ask each participant to estimate their relative performance compared to four randomly chosen participants present in the lab. Specifically, we ask each participant to indicate their best estimates (in percentage) that exactly 0, 1, 2, 3 or 4 of the other participants correctly answered more problems in the previous 5 minutes performance stage. We monetarily incentivize this belief elicitation using a quadratic scoring rule (Offerman 1997), which makes a participant’s earning higher the more accurate a participant’s estimates are (see Appendix 3.A.1 for details).

▪ **Second stage performance.** In this stage, we distinguish four conditions: three competitive conditions and a control condition. In each competitive condition, we randomly form mixed-sex pairs to compete against each other, and make participants aware of it. Then, we ask participants to perform the multiplication task for 5 minutes under a *winner-take-all tournament* incentive scheme. The instructions participants read explain the payment scheme as follows: “In this stage, you have been randomly paired with a participant of the opposite sex. That is, if you are a male participant, you are paired with a female participant. If you are a female participant, you are paired with a male participant. In this stage, you have to compete with your paired participant in order to earn money. The competition works as follows: the participant in your pair who correctly solves more multiplication problems earns € 0.40 for each correctly solved multiplication. The other earns nothing. In case you and the other participant correctly solve the same number of multiplication problems, then each of you earns € 0.20 for each correctly solved multiplication. Both of you see the same sequence of multiplication problems.”

The competitive payment scheme that we use in the second stage implies that being paid not only depends on the performance of others but also becomes uncertain. To minimize effects coming from differences in perceived risk in each

stage, we use a payment scheme in the non-competitive stage—the random pay—that makes payment uncertain, but not dependent on the performance of others.⁸

The only difference between the three competitive conditions lies in how we induce a stereotype threat.

Condition 1: Implicit stereotype threat (ST) against women. In this condition, we merely provide the instructions for the multiplication task. According to the proponents of the idea of implicit activation of stereotype threat (e.g., Smith and White 2002), the use of a mathematical task in our experimental setting is a sufficient situational clue to induce a woman to experience stereotype threat because the explicit mixed-sex pair formation makes sex salient, i.e., women are made aware that they are being compared to men.

Condition 2: Explicit ST against women. In this condition, we reinforce the stereotype threat against women embodied in the multiplication task by explicitly inducing a negative stereotype about women's ability to perform a mathematical task. Just before starting to compete, each participant unexpectedly faces the following information during 40 seconds:

“Before starting the multiplication task in this stage, please read carefully the following information:

In order to assess the magnitude of gender differences in mathematics performance, three leading researchers, J. Hyde, E. Fennema and J. Lamon, performed an evaluation of 100 studies in this field. In their paper we can read the following: “refined discussions generally conclude that the overall differences in mathematics performance (...) appear in adolescence and usually favor boys in tasks involving problem solving.”

In “Gender Differences in Mathematics Performance: A Meta-Analysis”, Psychological Bulletin.

If you wish, you can inspect this paper after the experiment.”

Condition 3: Explicit ST against men. In this condition, we contradict the stereotype that men are better at mathematics by explicitly inducing a negative stereotype about men's ability to perform a mathematical task. Just before starting to compete, each participant unexpectedly faces the following information during 40 seconds:

“Before starting the multiplication task in this stage, please read carefully the following information:

In order to assess the magnitude of gender differences in mathematics performance, three leading researchers, J. Hyde, E. Fennema and J. Lamon, performed an evaluation of 100 studies in this

⁸ If a standard piece-rate were used instead as the non-competitive payment scheme, gender differences in attitude towards risk that may obscure the pure competition effect on performance would be more likely to kick in.

field. In their paper we can read the following: "refined discussions generally conclude that the overall differences in mathematics performance (...) appear in adolescence and usually favor girls in tasks of computation."

In "Gender Differences in Mathematics Performance: A Meta-Analysis", Psychological Bulletin.

If you wish, you can inspect this paper after the experiment."

The quotes in Condition 2 and Condition 3 are not perfectly symmetric because we wanted to avoid deceiving participants.⁹ Thus, rather than fabricating symmetric quotes, we use truthful information that allows us to implement the desired manipulations while still minimizing text asymmetry.¹⁰

Finally, we conduct a control condition to assess whether there are any significant learning or fatigue effects during the performance of the multiplication task.

Condition 4: Twice random pay. In the second stage of this condition, we ask participants to perform the multiplication task for 5 minutes again under the random pay incentive scheme.

▪ **Risk attitude elicitation.** We elicit each participant response to the Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2011) general risk question: "How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, in which the value 0 means 'not at all willing to take risks' and the value 10 means 'very willing to take risks' ".¹¹

▪ **Competitive attitude elicitation.** We elicit each participant response to the Machiavelli personality test, also known as the Mach IV test (see Appendix 3.A.1 for details), in which high scores reliably predict competitive behavior (Christie and Geis 1970).

3.2.4. Experimental procedure

We recruit the participants with the help of ORSEE software (Greiner 2015) from the subject pool of the Behavioral and Experimental Economics Laboratory (BEElab) at Maastricht University. The lab experiment is computerized using z-Tree software (Fischbacher 2007) and conducted in the BEElab. We present all instructions on-screen and we treat all decisions confidentially. We conduct eight sessions (two sessions in each of the four conditions). In total, 192 subjects participate in the experiment (40, 46, 50 and 56 in the Implicit ST against women

⁹ A perfectly symmetric manipulation would be to use exactly the same text in both conditions, except for using the word "boys" in Condition 2 and replace it by the word "girls" in Condition 3.

¹⁰ The text pieces that we use are extracted from p. 140 of Hyde, Fennema, and Lamon (1990).

¹¹ In a paid field experiment Dohmen et al. (2011) show that responses to this question reliably predict incentivized lottery choices.

condition, Explicit ST against women condition, Explicit ST against men condition, and Twice random pay condition, respectively). An equal number of men and women participate in each session (we use the same procedure described in Chapter 2 to ensure a perfectly balanced gender composition in each session). The participants are predominantly students of Business and Economics at Maastricht University.¹² The average age is 22. Each session lasts on average 90 minutes. Average earnings are € 16.30.

3.3. Hypotheses

The main measure of the experiment is the change in performance between stages. The first stage is identical across conditions. In contrast, the second stage is different in each condition. In Table 3.1, we summarize the payment scheme that we use and the stereotype threat that we induce in each stage, per condition.¹³

Table 3.1. Payment scheme and induced stereotype threat per condition

Note: **IST-W** refers to the Implicit ST against women condition; **EST-W** refers to the Explicit ST against women condition; **EST-M** refers to the Explicit ST against men condition; **TRP** refers to the Twice random pay condition.

	1 st stage		2 nd stage	
	Payment Scheme	Stereotype Threat	Payment Scheme	Stereotype Threat
Condition IST-W	Random Pay	None	Competition	Implicit Against Women
Condition EST-W	Random Pay	None	Competition	Explicit Against Women
Condition EST-M	Random Pay	None	Competition	Explicit Against Men
Condition TRP	Random Pay	None	Random Pay	None

The control condition consists of two identical stages in which participants perform under a non-competitive payment. Based on Dohmen and Falk (2011), we do not expect any significant learning or fatigue effects during the performance of the multiplication task. Hence, we straightforwardly hypothesize:

Hypothesis 1: *The number of correct answers for men (women) in each stage is virtually the same in the Twice random pay condition. In other words, men's (women's) change in performance between stages is not significant.*

The winner-take-all tournament payment scheme that we introduce in the second stage is the same in each of the other three conditions. We expect this competitive incentive to induce higher effort and, as a result, improve the

¹² 86% of the participants are business and economics students. The remaining participants have a background in psychology, law, arts and culture, health sciences, European studies, or philosophy.

¹³ The first stage does not entail any of the triggers documented in the literature for an implicit activation of stereotype threat (which are: i. Make a participant's group membership salient; ii. Associate the real-effort task with a diagnostic of a participant's ability; iii. Fear of relative feedback or other's evaluation).

performance of participants who are not the target of a stereotype threat in the second stage. We therefore hypothesize:

Hypothesis 2: *Men in the Implicit and Explicit ST against women conditions as well as women in the Explicit ST against men condition significantly increase the number of correct answers in the competitive second stage.*

Women in the Implicit and Explicit ST against women conditions, and men in the Explicit ST against men condition are the target of a stereotype threat in the second stage. Simply put, stereotype threat theory implies that the performance of individuals belonging to a group is harmed in a situation entailing a stereotype threat targeting that group relatively to a situation entailing no stereotype threat (e.g., Steele and Aronson 1995). Therefore, the adverse effect of stereotype threat should counteract the expected positive effect of the competitive incentives on performance.

To make more detailed predictions according to stereotype threat theory, we measure two key elements: i. Verify that men and women from our subject pool do hold the stereotype that men are better at mathematics; ii. How men and women from our subject pool update their initial belief after reading the pieces of information that we use in the lab experiment to explicitly induce stereotype threats.

For this purpose, we conduct a monetarily incentivized online questionnaire. In the first part of the online questionnaire, we ask respondents to estimate the average number of correct answers by the male and female participants in the lab experiment during the 5 minutes non-competitive performance. The second part of the online questionnaire splits into two versions. In version I, half of the male and female respondents read the piece of information that we use in the lab experiment to explicitly induce a stereotype threat against women. In version II, the remaining male and female respondents read the piece of information that we use in the lab experiment to explicitly induce a stereotype threat against men. Subsequently, in both versions, we ask respondents whether they want to change their first estimates (see Appendix 3.A.2 for the full instructions, which provide the details of the monetary incentives). Importantly, the respondents to the online questionnaire and the participants in the lab experiment were distinct, but recruited from the same subject pool. Therefore, we are confident that the beliefs we elicit in the online questionnaire are representative of the beliefs of the participants in the lab experiment.¹⁴

¹⁴ The respondents to the online questionnaires share important demographics such as age and field of study with the participants in the lab experiment.

**Table 3.2. Average estimate of the gender gap in performance
(in number of correct answers)**

Note: A respondent's estimate of the gender gap in performance equals a respondent's estimate of correct answers of men minus a respondent's estimate of correct answers of women. ST-W (ST-M) refers to the piece of information that we use to explicitly induce a stereotype threat against women (men). ***, **, and * significant at 1%, 5%, and 10% derived from Wilcoxon signed-rank tests for the null hypothesis of a gender gap in performance equal to zero. Standard deviation in parentheses.

	Online Questionnaire			
	Version I (n=60)		Version II (n=60)	
	Before info	After ST-W info	Before info	After ST-M info
Male respondents	2.2** (6.4)	5.8*** (5.2)	1.9** (5.8)	-2.1*** (4.5)
Female respondents	2.8** (6.7)	4.9*** (6.4)	2.7*** (4.7)	-1.4* (4.3)

In Table 3.2, we report male and female respondents' estimate of the gender gap in performance. We see that, before receiving the stereotyping piece of information, male and female respondents in both versions of the online questionnaire believe that men perform the multiplication task significantly better than women. In other words, these data is consistent with men and women from our subject pool holding the stereotype that men are better at mathematics. Moreover, with reference to version I, we observe that male and female respondents' initial belief in a gender gap in performance favoring men is amplified after we explicitly induce a stereotype threat against women. Finally, with reference to version II, we observe that the male and female respondents reverse their initial belief in a gender gap in performance favoring men after we explicitly induce a stereotype threat against men. Thus, our third and last hypothesis is:

Hypothesis 3: *The change in performance between stages of women is significantly weaker than men's in the Implicit and Explicit ST against women conditions. The change in performance between stages of men is significantly weaker than women's in the Explicit ST against men condition.*¹⁵

¹⁵ The hypotheses we present in this section are consistent with the model of stereotypes recently proposed by Bordalo, Gennaioli, and Schleifer (2014).

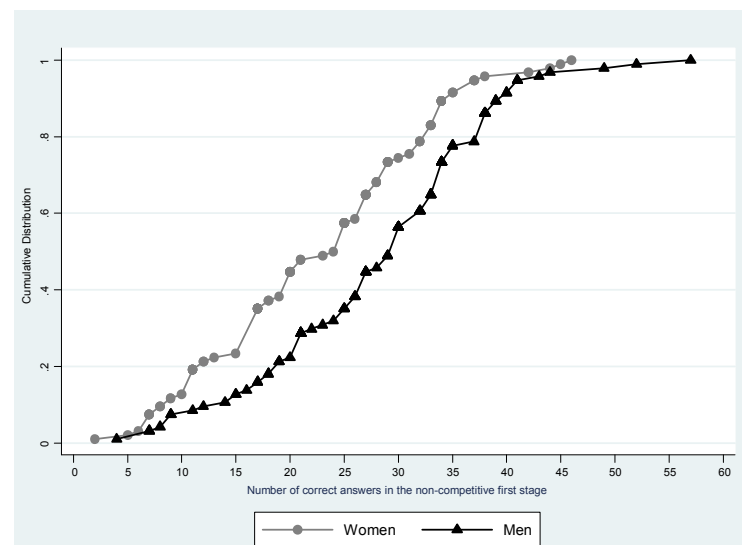
3.4. Results of the lab experiment

Throughout this section, we report either Wilcoxon signed-rank (WSR) tests or Mann-Whitney (MW) tests to investigate differences in distributions. All tests we report are two-sided.

3.4.1. Non-competitive performance vs. competitive performance

In Figure 3.2, we plot the cumulative distribution of the number of correct answers during the non-competitive first stage performance, pooling the four conditions. We observe that the men's performance distribution statistically dominates women's performance distribution, i.e., men perform significantly better than women.¹⁶ Moreover, we also observe in Figure 3.2 a large inter-individual heterogeneity in performance. This aspect does not impede our analysis because our main measure—the within-subject change of performance from the non-competitive to the competitive stage—accounts for the heterogeneity in baseline performance.

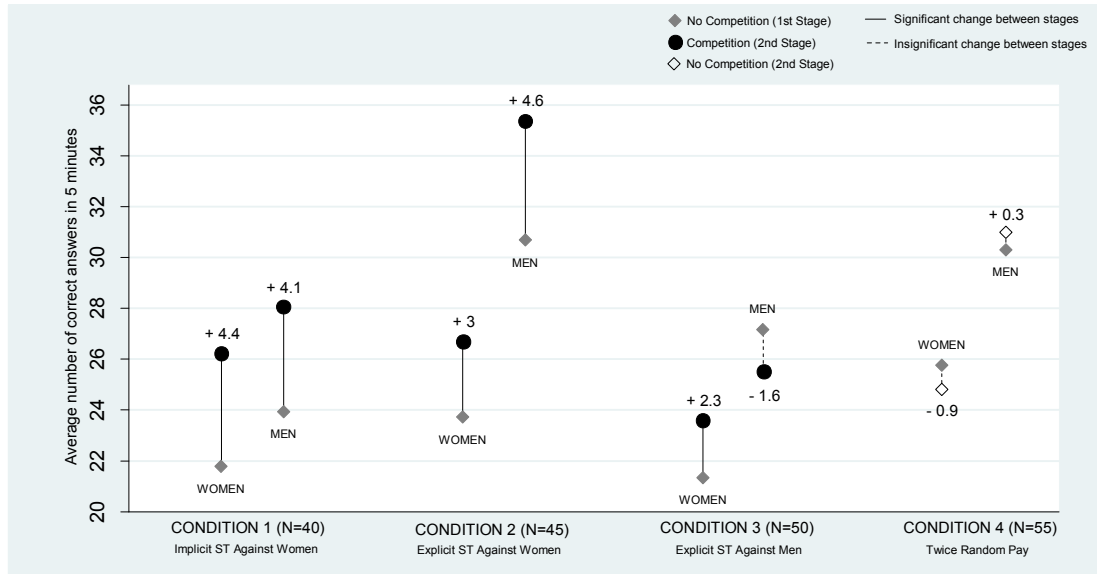
Figure 3.2. Cumulative distributions of the non-competitive first stage performance (pooling all conditions)



¹⁶ The pooled results for men's and women's average number of correct answers in the first stage are: 28.3 (men) vs. 23.2 (women), $p < 0.001$, MW test, $n = 190$. Broken down by condition, the figures are: i. Condition IST-W: 23.9 (men) vs. 21.8 (women), $p = 0.507$, MW test, $n = 40$. ii. Condition EST-W: 30.8 (men) vs. 23.7 (women), $p = 0.024$, MW test, $n = 45$. iii. Condition EST-M: 27.2 (men) vs. 21.3 (women), $p = 0.024$, MW test, $n = 50$. iv. Condition TRP: 30.4 (men) vs. 25.7 (women), $p = 0.175$, MW test, $n = 55$. We do not include throughout the data analysis the behavior of two male participants who participated in condition EST-W and TRP, respectively. In both stages, these participants hit the number of problems available (70) before time elapsed. The results we present are qualitatively the same if we include these two participants.

To analyze how men and women respond to competitive incentives in a *mixed-sex competition* we make two types of comparison. Firstly, we compare men's (women's) performance in the non-competitive stage to their performance in the competitive stage. Secondly, to evaluate possible gender differences in the response to competition, we perform a difference in differences analysis. That is, we compare men's to women's change in performance between stages.

Figure 3.3. Change in performance between stages



In Figure 3.3, we summarize the performance results in each condition. To check whether we have to take into account learning or fatigue effects, we first analyze the Twice random pay condition. In this condition, the average number of correct answers by men is 30.4 (SD: 10.35) in the first stage and 30.7 (SD: 10.13) in the second stage. The difference is not significant ($p = 0.819$, WSR test). Women correctly answer on average 25.7 (SD: 11.58) problems in the first stage and 24.8 (SD: 15.33) in the second stage. This difference is not significant either ($p = 0.515$, WSR test). In addition, we find for both sexes a substantial overlap between the cumulative distributions of the first and second stage performance (see Appendix 3.A.3), which corroborates the finding that there is neither a fatigue nor an experience effect for both sexes. Therefore, our first result is:

Result 1: *In the Twice random pay condition, men's and women's change in performance between stages is not significant.*

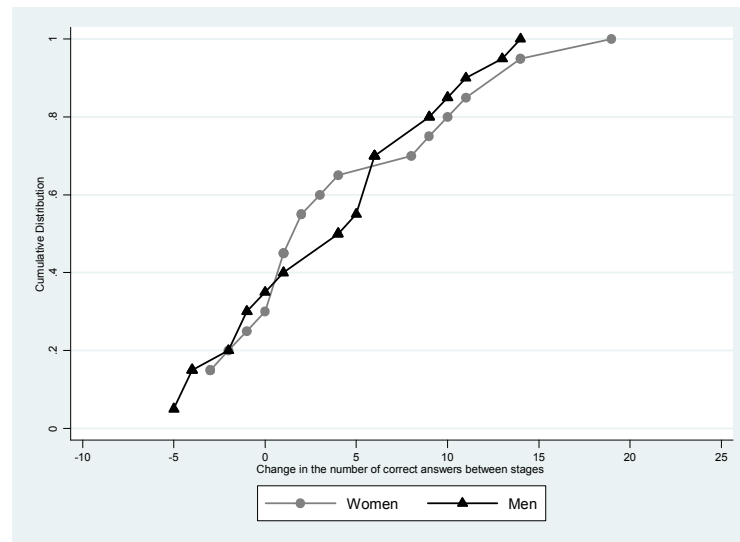
Hence, throughout our analysis we consider that men's and women's performance of the multiplication task in the other conditions is not affected by repetition either.

In the Implicit ST against women condition, the average number of correct answers by men is 23.9 (SD: 11.54) in the non-competitive first stage and 28 (SD:

13.23) in the competitive second stage. This difference is significant ($p = 0.012$, WSR test). Women correctly answer on average 21.8 (SD: 10.79) problems in the non-competitive stage and 26.2 (SD: 11.25) problems in the competitive stage. This difference is also significant ($p = 0.019$, WSR test). Moreover, the cumulative distributions of performance in the first and second stages show that both men's and women's performance increases under competition (see Appendix 3.A.3).

Comparing men's to women's increase in performance between stages reveals that they are not significantly different [$+ 4.1$ (men) vs. $+ 4.4$ (women), $p = 0.968$, MW test]. The proximity of the cumulative distributions in Figure 3.4 illustrates the latter finding. We summarize in our second result.

Figure 3.4. Cumulative distributions of the change in performance between stages (Implicit ST against women condition)

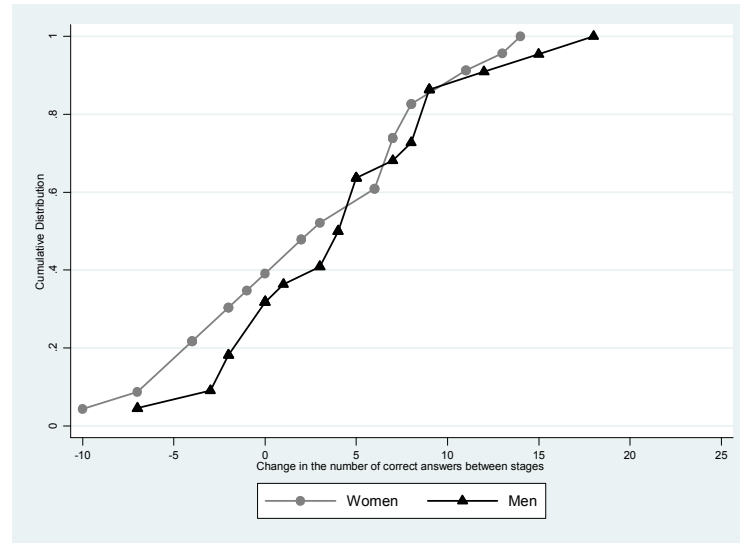


Result 2: *When there is an implicit ST against women, men and women significantly increase their performance under competition. Notably, their increase in performance between stages is similar.*

In the Explicit ST against women condition, men correctly answer on average 30.8 (SD: 10.10) problems when there is no competition and 35.4 (SD: 12.97) if there is competition. This difference is significant ($p = 0.004$, WSR test). Women's average performance in this condition is 23.7 (SD: 9.71) and 26.7 (SD: 11.12) correct answers in the non-competitive and the competitive stage, respectively. This difference is also significant ($p = 0.044$, WSR test). Again, the cumulative distributions of performance in the first and second stages show that both men's and women's performance increases under competition (see Appendix 3.A.3).

Comparing men's to women's increase in performance between stages, we find that they are statistically indistinguishable [$+ 4.6$ (men) vs. $+ 3$ (women), $p = 0.510$,

Figure 3.5. Cumulative distributions of the change in performance between stages (Explicit ST against women condition)



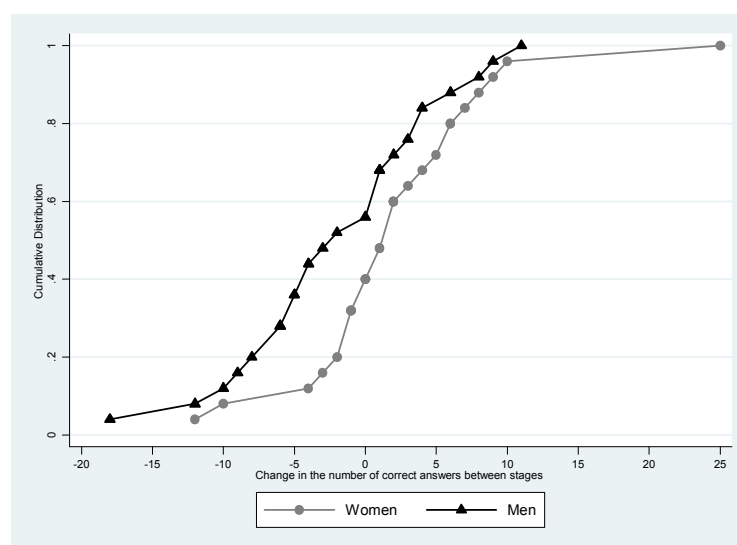
MW test]. The proximity of the cumulative distributions in Figure 3.5 illustrates the latter finding. We sum up in our third result.

Result 3: *When there is an explicit ST against women, men and women significantly increase their performance under competition. Again, their increase in performance between stages is similar.*

In the Explicit ST against men condition, the picture is different. In this condition, the average number of correct answers by men is 27.2 (SD: 9.45) in the non-competitive stage and 25.6 (SD: 10.33) in the competitive stage. This means men perform worse under competition than under no competition, but the difference is not significant ($p = 0.269$, WSR test). Women perform better under competition than under no competition, with 23.6 (SD: 11.38) and 21.3 (SD: 8.84) correct answers, respectively. This difference is marginally significant ($p = 0.090$, WSR test). The men's cumulative distributions of competitive and non-competitive performance confirm that men's performance did not increase under competition. In respect to women, the analogous cumulative distributions support that they increase performance under competition (see Appendix 3.A.3).

Comparing men's to women's change in performance between stages, we find a significant gender difference in response to competitive incentives [-1.6 (men) vs. $+2.3$ (women), $p = 0.066$, MW test]. In Figure 3.6, we corroborate this finding: women's change in performance distribution statistically dominates men's change in performance distribution. We summarize the findings in this condition in Result 4.

Figure 3.6. Cumulative distributions of the change in performance between stages (Explicit ST against men condition)



Result 4: *When there is an explicit ST against men, women improve performance between stages whereas men do not.*

To further evaluate the treatment effects on men's and women's competitive performance, we run a regression analysis. In Table 3.3, we present three linear regression models of change in performance between stages. In regressions (1) and (3), the base group is men in the Twice random pay condition. In regression (2), the base group is women in the Twice random pay condition.

The results in regressions (1) and (2) are in line with the previous statistical and graphical analysis. In each regression, the intercept is not significant, which means that the performance between stages is not affected by repetition in the Twice random pay condition for men and women, respectively.¹⁷ Moreover, considering the Implicit ST against women condition, we observe that men (with 3.293 more correct answers) and women (with 5.335 more correct answers) significantly improve their performance between stages relatively to men and women in the Twice random pay condition, respectively. The same holds for men (with 4.278 more correct answers) and women (with 4.011 more correct answers) in the Explicit ST against women condition. With respect to the Explicit ST against men condition, men's change in performance between stages (with 2.201 less correct answers) is worse relatively to men in the Twice random pay condition, but the difference is not significant. In contrast, women's change in performance between stages (with 3.282 more correct answers) is significantly better relatively to women in the Twice random pay condition.

¹⁷ The intercept measures the change in the number of correct answers between stages by the participants in the base group with a non-competitive performance equal to the sample mean.

Table 3.3. Linear regression model of change in performance between stages I

Note: We demean the non-dummy regressors so that the interpretation of the intercept is meaningful (e.g., $Non-competitive\ performance_i = \text{actual non-competitive performance}_i - \text{sample mean of non-competitive performance}$). ***, **, and * significant at 1%, 5%, and 10%, respectively. Robust standard errors in brackets.

Dependent variable:	change in performance between stages		
	Men (1)	Women (2)	Pooled (3)
<i>Non-competitive performance</i>	-0.071 [0.065]	0.033 [0.060]	-0.018 [0.045]
<i>1 if Implicit ST against women</i>	3.293* [1.969]	5.335*** [1.938]	3.634* [1.928]
<i>1 if Explicit ST against women</i>	4.278** [1.871]	4.011** [1.877]	4.256** [1.865]
<i>1 if Explicit ST against men</i>	-2.201 [1.944]	3.282* [1.857]	-2.035 [1.951]
<i>1 if female</i>	n.a.	n.a.	-1.241 [1.865]
<i>female x Implicit ST against women</i>	n.a.	n.a.	1.502 [2.701]
<i>female x Explicit ST against women</i>	n.a.	n.a.	-0.350 [2.641]
<i>female x Explicit ST against men</i>	n.a.	n.a.	5.091* [2.701]
<i>Intercept</i>	0.447 [1.325]	-0.938 [1.263]	0.383 [1.369]
Observations	94	96	190
R ²	0.151	0.084	0.112

In regression (3), we pool the data to directly compare men's to women's response to competition. We observe that the interaction terms *female x Implicit ST against women* and *female x Explicit ST against women* are insignificant. This means that women's change in performance between stages is not significantly different to men's both in the Implicit and the Explicit ST against women conditions. In short, men's and women's reaction to competitive incentives is similar in these two conditions. Differently, in the Explicit ST against men condition we see that the interaction term *female x Explicit ST against men* is significant, which means that women's competitive response is significantly stronger than men's.

In a nutshell, the regression analysis corroborates Results 1 to 4.

3.4.2. Confidence level, risk attitude, and competitive attitude

The payment each participant faces in the competitive second stage depends not only on their performance but also on another participant's performance. Therefore, participants' belief about their relative rank might affect participants' performance under competition.

In Table 3.4, we report that men estimate, on average, that they are the best performer in 19.62% of the case, which is significantly higher than the corresponding women's estimate of 8.65%. In contrast, men estimate, on average, that they are the worst performer in 7.65% of the cases, which is significantly lower than the corresponding women's estimate of 21.29%. The gender differences in the remaining rank estimates are statistically insignificant.

Table 3.4. Self-assessed performance rank (in average percentage)

Note: Recall that right after the non-competitive first stage, we ask each participant to indicate their best estimates that exactly 0 (rank 1), 1 (rank 2), 2 (rank 3), 3 (rank 4) or 4 (rank 5) out of four randomly chosen participants in the experimental session correctly answered more problems than the participant in the non-competitive first stage. p-values derived from Mann-Whitney tests of equality of distributions between men and women.

	Rank				
	1: Best	2	3	4	5: Worst
Men	19.62	22.56	28.18	21.99	7.65
Women	8.65	20.97	24.24	24.85	21.29
p-value	0.001	0.145	0.141	0.604	0.003

To further assess the confidence level, we compute for each participant a confidence index that aggregates all the self-assessed confidence beliefs.¹⁸ According to our confidence index, men's and women's average rank estimate is 2.75 (SD: 0.90) and 3.29 (SD: 0.98), respectively. This difference is highly significant ($p < 0.001$, MW test). Therefore, we conclude that men are more confident than women to perform the multiplication task just before the competitive second stage.¹⁹

$$^{18} \left[\sum_{i=0}^4 p_i \times (i+1) = p_0 \times 1 + p_1 \times 2 + p_2 \times 3 + p_3 \times 4 + p_4 \times 5 \right] / 100 \text{ yields a participant's}$$

confidence index, in which i represents the outcome that exactly i other participants correctly answered more problems and p_i is the participant's percentage estimate that outcome i is the actual one. Note that a lower index indicates a more confident participant.

¹⁹ This finding is not driven by men performing better the multiplication task. In unreported regressions, we find that men are significantly more confident than women when controlling for the non-competitive performance.

Another factor that might cause gender differences in performance between the non-competitive and the competitive stage is a difference in men's and women's risk attitude. The average response of men to the Dohmen et al. (2011) general risk question, 5.95 (SD: 2.14), is significantly higher than women's average response, 5.09 (SD: 2.10) ($p = 0.007$, MW test). These results are in line with the evidence that women are, on average, more risk averse than men (e.g., Croson and Gneezy, 2009).

Finally, we elicit participants' competitive attitude. We assume that a more competitive person puts more effort in the multiplication task during the competitive stage. We should therefore expect a higher performance increase on such a person. The average score of men in the Mach-IV test is with 79.52 (SD: 12.64) significantly higher than women's average score of 74.30 (SD: 12.41), ($p = 0.002$, MW test). Hence, according to the Mach-IV test, men are significantly more competitive than women.²⁰

To determine how these factors jointly affect the change in performance between stages, and to understand their relative significance, we report in Table 3.5 augmented versions of the models presented in Table 3.3, which add confidence index, risk attitude, and competitive attitude to the set of explanatory variables. In regressions (1) and (3), the base group is men in the Twice random pay condition. In regression (2), the base group is women in the Twice random pay condition.

The results in regressions (1) and (2) of Table 3.5 show that the significance (insignificance) results that we identified in regression (1) and (2) of Table 3.3 are robust when controlling for the additional explanatory variables. The same holds when we directly compare men's to women's response to competition based on the estimates in regression (3) of Table 3.5. These observations lead to our next result.

Result 5: *The treatment effects regarding the change in performance between stages are robust when controlling for confidence level, risk attitude, and competitive attitude.*

²⁰ It is reasonable to interpret the Mach IV test as only measuring competitive attitude with "elbows". Thus, a possible limitation of this test is that it does not elicit other traits of competitive individuals.

Table 3.5. Linear regression model of change in performance between stages II

Note: We demean the non-dummy regressors so that the interpretation of the intercept is meaningful (e.g., $Non-competitive\ performance_i = \text{actual non-competitive performance}_i - \text{sample mean of non-competitive performance}$). ***, **, and * significant at 1%, 5%, and 10%, respectively. Robust standard errors in brackets.

Dependent variable:	<i>change in performance between stages</i>		
	Men (1)	Women (2)	Pooled (3)
<i>Non-competitive performance</i>	-0.066 [0.089]	0.096 [0.059]	0.018 [0.051]
<i>1 if Implicit ST against women</i>	3.087* [1.814]	5.508*** [2.011]	3.731** [1.817]
<i>1 if Explicit ST against women</i>	4.184** [1.822]	4.159** [2.013]	4.439** [1.813]
<i>1 if Explicit ST against men</i>	-1.890 [2.074]	3.262* [1.806]	-1.431 [2.006]
<i>Confidence level</i>	-0.617 [0.995]	1.458* [0.749]	0.617 [0.606]
<i>Risk attitude</i>	0.780*** [0.298]	0.162 [0.369]	0.476** [0.240]
<i>Competitive attitude</i>	0.103* [0.052]	0.105 [0.072]	0.098** [0.046]
<i>1 if female</i>	n.a.	n.a.	-0.664 [1.827]
<i>female x Implicit ST against women</i>	n.a.	n.a.	1.962 [2.642]
<i>female x Explicit ST against women</i>	n.a.	n.a.	0.137 [2.628]
<i>female x Explicit ST against men</i>	n.a.	n.a.	4.985* [2.672]
<i>Intercept</i>	0.430 [1.274]	-1.004 [1.362]	-0.225 [1.343]
Observations	94	96	190
R ²	0.234	0.154	0.167

3.4.3. Is there a stereotype threat shadow? A direct test

Consider two groups that simultaneously perform a task in two distinct situations. In one situation, one of the groups is negatively stereotyped relatively to the other. In the other situation, neither group is exposed to stereotype threat. Stereotype threat theory predicts that under such circumstances the difference in performance between the two groups is not the same in each situation because the performance of the targeted group is harmed in the situation entailing a stereotype threat.

In the first stage of the lab experiment, in which neither men nor women are exposed to stereotype threat,²¹ we observe that the gender gap in performance is favorable to men in each condition. Therefore, stereotype threat theory would predict: i. The gender gap in performance increases in the second stage of the Implicit and the Explicit ST against women conditions; ii. The gender gap in performance decreases in the second stage of the Explicit ST against men condition; iii. The gender gap in performance keeps virtually the same in the second stage of the Twice random pay condition. To test these hypotheses, we compare in each condition the gender gap in performance in the second stage to the gender gap in performance in the first stage.

In Figure 3.7, we display the average gap in performance between men and women in each stage, per condition.²² In the Twice random pay condition, in which we do not induce a stereotype threat in any stage, the average gap in performance between men and women is 4.9 (SD: 16.29) correct answers in the first stage and 6 (SD: 19.62) correct answers in the second stage. As expected, the difference between stages of the gender gap in performance is not significant (4.9 vs. 6, $p = 0.580$, WSR test).

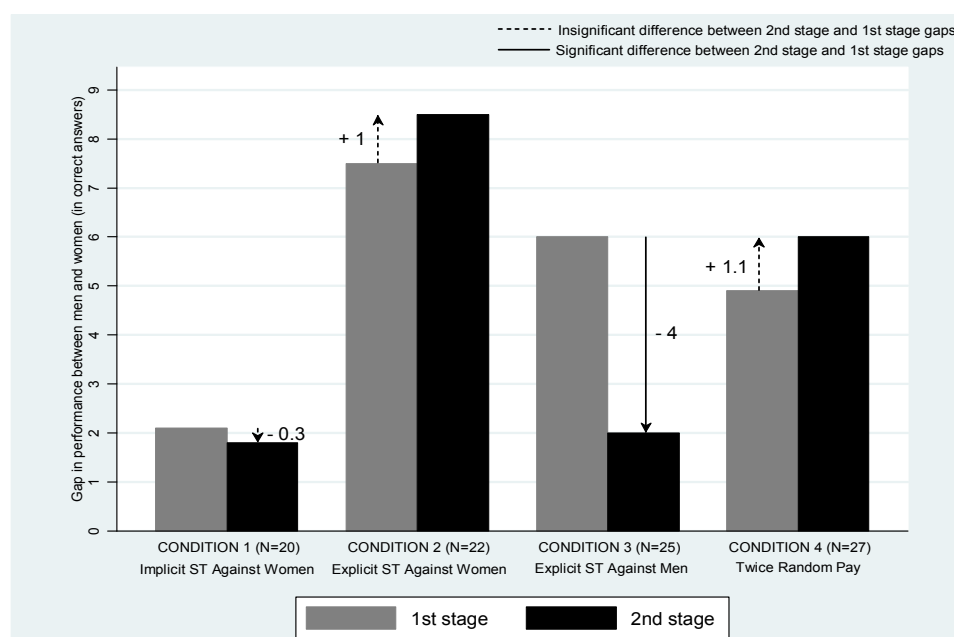
In the Implicit ST against women condition, the average gap in performance between men and women is 2.1 (SD: 13.70) correct answers in the first stage and 1.8 (SD: 15.41) correct answers in the second stage. The decrease of the gender gap in performance in the second stage is not significant (2.1 vs. 1.8, $p = 0.970$, WSR test). Stereotype threat theory would predict in this case an increase of the gender gap in performance in the second stage.

In the Explicit ST against women condition, the average gap in performance between men and women is 7.5 (SD: 12.42) correct answers and 8.5 (SD: 17.99)

²¹ See footnote 13.

²² The average gender gap in performance in the 2nd stage equals $\sum_{i=1}^n (2^{nd} \text{ stage performance male}_i - 2^{nd} \text{ stage performance female}_i) / n$, in which n is the total number of pairs that we form for the competition in the second stage. The average gender gap in performance in the 1st stage equals $\sum_{i=1}^n (1^{st} \text{ stage performance male}_i - 1^{st} \text{ stage performance female}_i) / n$, in which we keep the pairs that we form for the second stage. To compute the gender gaps in the Twice random pay condition, we use in both stages the pairs that we form to determine who gets paid in the second stage.

Figure 3.7. Gap in performance between men and women in each stage



correct answers in the first and second stage, respectively. The increase of the gender gap in performance in the second stage is consistent with stereotype threat theory, but the difference between the two stages is not significant (7.5 vs. 8.5, $p = 0.603$, WSR test). Therefore, we conclude that there is no evidence for the adverse effect of stereotype threat in this condition either.

Finally, in the Explicit ST against men condition, the average gap in performance between men and women is 6 (SD: 13.21) correct answers in the no threat first stage and 2 (SD: 14.83) correct answers in the second stage. The gender gap in performance in the second stage is significantly lower than in the first stage (6 vs. 2, $p = 0.082$, WSR test), which is consistent with an explanation based on stereotype threat.

These results do not support the common finding that women's performance is impaired in situations entailing a mathematical stereotype threat against women. Notwithstanding, we find evidence for the adverse effect of stereotype threat when we explicitly induce a stereotype threat against men. The relative worsening of men's performance in our setting when informed about women's superior ability in mathematics is significant, even though men are not considered a group prone to a mathematical stereotype threat.²³ We summarize in the next result.

Result 6: *Solely the treatment effect regarding men's and women's performance in the Explicit ST against men condition are in line with an explanation based on stereotype threat.*

²³ According to stereotype theory, performance worsening should be particularly observed if the targeted group is prone to the stereotype threat. However, proneness is not a necessary condition to experience stereotype threat. For instance, Aronson et al. (1999) find that Caucasian men perform worse in math-tests when informed about male Asian-Americans' superior ability in mathematics.

3.4.4. What causes change in performance?

In this section, we analyze men's and women's performance in more detail. We assume that a necessary condition for each participant to improve performance between stages is to increase effort in the competitive second stage. Hence, if a participant increases the number of correct answers between stages, we should be able to detect one of the following three forms of increase in effort during the second stage: i. Increase in the number of attempts to solve problems holding constant the accuracy rate of the first stage; ii. Increase in the accuracy rate holding constant the number of attempts of the first stage; iii. Increase in both the number of attempts and the accuracy rate.²⁴

In Table 3.6, we report participants' number of attempts to solve problems²⁵ and their accuracy rate. As expected, men's and women's number of attempts and

Table 3.6. Attempts and accuracy rate

Note: A participant's accuracy rate equals the number of correct answers divided by the number of attempts to solve problems. We do not include two female participants in the Twice random pay condition because they did not attempt any question in the second stage (we interpret their behavior as indicative of boredom). *** and ** significant at 1% and 5%, respectively, derived from Wilcoxon signed-rank tests.

		Average number of attempts			Average accuracy rate		
		1 st stage	2 nd stage	Difference	1 st stage	2 nd stage	Difference
Implicit ST against women	Women	28.55	33.05	4.5***	0.737	0.779	0.042
	Men	29.4	32.95	3.55***	0.802	0.841	0.039
Explicit ST against women	Women	29.35	32.13	2.78***	0.795	0.823	0.028
	Men	35.86	41.55	5.69***	0.848	0.836	-0.012
Explicit ST against men	Women	26.76	29.64	2.88***	0.752	0.758	0.006
	Men	31.72	31.80	0.08	0.850	0.785	-0.065***
Twice random pay	Women	31.35	31.65	0.30	0.851	0.821	-0.030
	Men	36.44	37.07	0.63	0.817	0.822	0.004

²⁴ The following examples illustrate the three forms we consider. Assume a participant makes 4 attempts during the first stage performance that yield 2 correct answers and 2 errors. E.g. form 1: The participant makes 8 attempts during the second stage performance that yield 4 correct answers and 4 errors. E.g. form 2: The participant makes 4 attempts during the second stage performance that yield 4 correct answers. E.g. form 3: The participant makes 8 attempts during the second stage performance that yield 8 correct answers.

²⁵ The z-Tree code counts an attempt each time a participant provides a numerical answer. Recall that if a wrong answer is provided, a participant has to attempt the same problem again.

accuracy rate is not significantly different between stages in the Twice random pay condition. Regarding the conditions in which we induce stereotype threat, we observe that men and women in the Implicit and Explicit ST against women conditions as well as women in the Explicit ST against men condition significantly increase the number of attempts in the second stage while not significantly changing the accuracy rate. These observations lead to our seventh result.

Result 7: *Men and women in the Implicit and Explicit ST against women conditions as well as women in the Explicit ST against men condition significantly improve performance under competition because they tackle more problems while keeping the same accuracy. In other words, they efficiently exert more effort under competition.*

Regarding men in the Explicit ST against men condition, we observe in Table 3.6 that they make virtually the same number of attempts in each stage and their accuracy rate significantly drops in the second stage.

Result 8: *Men in the Explicit ST against men condition fail to correctly answer more problems under competition because their accuracy rate decreases while virtually tackling the same number of problems.*

3.4.5. Discussion—Why do men underperform when under stereotype threat?

Relatively to the other five cases—men and women in the Implicit and the Explicit ST against women conditions and women in the Explicit ST against men condition—, we could be inclined to interpret men’s failure to increase the number of attempts in the Explicit ST against men condition as evidence that they do not exert more effort in the second stage. However, men’s significantly lower accuracy rate in the second stage rather suggests an explanation based on performance impairment.

In this section, we examine more closely the behavior of men under stereotype threat compared to men not under stereotype threat to see whether we may shed light on the effort provision of men in the Explicit ST against men condition. Specifically, we examine men’s change in performance between stages broken down in three periods of 100 seconds to test whether the poorer performance of men under stereotype threat relatively to men not under stereotype threat is persistent along the 5 minutes performance under competition.

In Table 3.7, we report men’s change in performance between stages broken down in three periods of 100 seconds. With reference to the last row of Table 3.7, we observe that the change in the accuracy rate of men under stereotype threat is weaker relatively to men under no stereotype threat in each period, but the effect is significant only in the initial 100 seconds of performance (-0.175). Moreover, we observe that the change in the number of attempts is significantly weaker in the first two periods, the effect is more pronounced in the second period (-2.11), and fades away in the last period. In fact, we cannot reject that in the Period III men

Table 3.7. Men's change in attempts and accuracy rate between stages (broken down in three periods)

Note: *Period I*, *Period II*, and *Period III* refer to the initial, middle, and final 100 seconds of the 5 minutes performance, respectively. *Men under ST* refers to men in the Explicit ST against men condition. *Men not under ST* refers to men in the Implicit and Explicit ST against women conditions. We pool the latter two conditions because men's change in the number of attempts between stages is not significantly different across periods in these conditions. The same holds for the change in the accuracy rate between stages. *** and ** significant at 1% and 5% level, respectively, derived from Mann-Whitney tests.

	Change in number of attempts between stages			Change in accuracy rate between stages		
	Period I	Period II	Period III	Period I	Period II	Period III
Men under ST	-0.88	-0.56	1.52	-0.148	-0.036	-0.040
Men not under ST	0.64	1.55	2.48	0.027	0.028	0.029
Difference	-1.52**	-2.11***	-0.96	-0.175***	-0.064	-0.069

under stereotype threat increase the number of attempts as much as men not under stereotype threat do (1.52 vs. 2.48).

We now look at Table 3.7 from a within-subject perspective. With reference to the first row of Table 3.7, we observe that the accuracy rate of men under stereotype threat noticeably improves in the middle 100 seconds of performance (Period I: -0.148 vs. Period II: -0.036, $p = 0.022$, WSR test), and keeps roughly the same in the last period (Period II: -0.036 vs. Period III: -0.040, $p = 0.778$, WSR test). Moreover, the change in the number of attempts is increasing across periods, and the positive 1.52 change in the last period is significant [1st stage (Period II): 10.56 vs. 2nd stage (Period III): 12.08, $p = 0.026$, WSR test].

In short, the performance of men under stereotype threat is poorer than men under no stereotype threat, but from a within-subject perspective they improve their performance along the 5 minutes competition. The latter finding makes it implausible to attribute men's poorer performance in the Explicit ST against men condition to lack of effort in the competitive stage. Despite men's performance in the Explicit ST against men is poorer from a between-subject perspective, the fact that from a within-subject perspective it improves along the competition indicates that men are trying their best in the second stage of the Explicit ST against men condition.

A possible explanation for the results is that men under stereotype threat do provide more effort in the second stage, but cannot convert the extra effort into correct answers in the initial phase of the competition because of performance impairment. That is, the adverse effect of stereotype threat causes men to make

too many errors in the initial period of the competition, which slows down their performance relatively to the first stage. As time goes on, men eventually manage to overcome the adverse effect of stereotype threat. As a result, men improve their accuracy rate and speed up their performance. In the final period of the competition, men under stereotype threat not only already perform faster than in the first stage (within-subject comparison) but also perform equally well as men in the other two conditions (between-subject comparison). The average change in performance between stages in Period III corroborates this interpretation: This score for men under stereotype threat in the last period is 1.2 more correct answers. The corresponding score for men under no stereotype threat is 2.3 more correct answers, which is not significantly different than 1.2 (2.3 vs. 1.2, $p = 0.438$, MW test).

However, note that the latter explanation is just a possible interpretation of our results based on suggestive evidence. Strictly speaking, our data do not allow to disentangle performance impairment from effort provision.

3.5. Effort provision—A normative view

In this section, we first describe the theoretical framework we use to investigate the effort level participants might exert in each performance stage of the lab experiment. Second, we concisely report the main conclusions. Finally, we relate the normative predictions to the results of the lab experiment.

3.5.1. Framework

Consider an arbitrary participant i of the lab experiment. We assume that participant i 's number of correct answers in each stage equals e_i , which represents the effort participant i exerts. In line with the lab experiment, the size of the monetary prize participant i may earn in each performance stage is not fixed ex-ante, but rather depends positively on effort provision. Specifically, participant i 's revenue is represented by $R_i(e_i) = 0.40 \cdot e_i$ in which 0.40 indicates participant i 's earnings per correct answer in Euro cents, $R'_i(e_i) > 0$ and $R''_i(e_i) = 0$. Therefore, we assume that the accuracy rate to convert effort into correct answers is the same in both stages. Moreover, participant i 's cost to exert effort is represented by a strictly convex function $c_i : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$. Specifically, we assume that $c_i(e_i) = \frac{\alpha_i}{2} \cdot e_i^2$ in which $\alpha_i > 0$ indicates participant i 's ability to correctly solve multiplication problems, $c_i(0) = 0$, $c'_i(e_i) > 0$ for $e_i > 0$ and $c''_i(e_i) > 0$.

3.5.1.1. First stage: performance under a random pay incentive scheme

Participant i faces the following problem in the first stage of the lab experiment:

$$\max_{e_i} \pi_i(e_i) \quad \text{with} \quad \pi_i(e_i) = \text{prob}^{rp} \cdot 0.4 \cdot e_i - \frac{\alpha_i}{2} \cdot e_i^2$$

in which $\text{prob}^{rp} = 0.5$ represents the probability of being paid under the random pay incentive scheme.

3.5.1.2. Second stage: performance under a winner-take-all tournament incentive scheme

The second stage of the lab experiment consists of a one-to-one mixed-sex competition in which participants perform the multiplication task during 5 minutes under a winner-take-all tournament incentive scheme. To analyze the strategic nature of this stage, we consider a static game of complete information.²⁶ Specifically, we consider a static game in which both players, to whom we generically refer to as participant i and participant j , have to simultaneously choose their effort level before the start of the competition. Thus, a possible outcome of the game is any pair of strategies (e_i^{com}, e_j^{com}) in which e_i^{com} and e_j^{com} are the effort levels that participant i and participant j , respectively, choose in the competitive second stage. Without loss of generality, we proceed the exposition considering participant i 's perspective.

In this competitive setting, participant i probability of winning the tournament against participant j is described by the following *Tournament Success Function*:

$$\text{prob}_i^{com}(e_i^{com}, e_j^{com}) = \begin{cases} 1 & \text{if } e_i^{com} \geq e_j^{com} \\ 0 & \text{if } e_i^{com} < e_j^{com} \end{cases}$$

in which prob_i^{com} represents the probability of participant i winning the competition against participant j , i.e., the probability of being paid under the winner-take-all tournament incentive scheme.²⁷

²⁶It would certainly be more realistic to consider a static game of incomplete information, in which participants are privately informed about their ability parameter a . In the Appendix 3.A.7, we concisely depict the economics literature on contests and tournaments to explain why we instead put forth a static game of complete information.

²⁷Strictly speaking, the participants in the lab experiment equally split the total earnings in case of a tie. In our analysis, we instead assume that both participants still get the total earnings in case of a tie. Nevertheless, we report the consequences for the Nash equilibria that we find if we did not make this simplifying assumption.

Hence, participant i faces the following problem in the second stage of the lab experiment:

$$\max_{e_i} \pi_i \left(e_i | b_{i,e_j^{com}} \right) \text{ with } \pi_i \left(e_i | b_{i,e_j^{com}} \right) = prob_i^{com} \left(e_i | b_{i,e_j^{com}} \right) \cdot 0.4 \cdot e_i - \frac{\alpha_i}{2} \cdot e_i^2$$

in which $b_{i,e_j^{com}}$ represents participant i 's belief about the effort participant j will exert in the competitive stage. Thus, $prob_i^{com} \left(e_i | b_{i,e_j^{com}} \right) = 1$ if $e_i^{com} \geq b_{i,e_j^{com}}$ or $prob_i^{com} \left(e_i | b_{i,e_j^{com}} \right) = 0$ if $e_i^{com} < b_{i,e_j^{com}}$.

To derive Nash equilibria in our competitive setting, we consider three cases: i. The *Symmetric case*, in which participant i 's ability equals participant j 's ($\alpha_i = \alpha_j$); ii. The *Asymmetric case I*, in which participant i 's ability is higher than participant j 's ($\alpha_i < \alpha_j$), but the difference in ability is not *considerable* (in the sense that both participants nonetheless believe that participant j still has a chance to win the competition); iii. The *Asymmetric case II*, in which participant i 's ability is *considerably* higher than participant j 's ($\alpha_i \ll \alpha_j$).

3.5.2. Effort under random pay vs. effort under competition

Essentially, we can summarize the conclusions of our normative analysis as follows (see Appendices 3.A.4, 3.A.5, and 3.A.6 for the detailed analysis, derivation of Nash equilibria, and proofs):

Conclusion 1. *In each Nash equilibrium of the Symmetric case, the participants exert more effort under competition than under random pay. The same holds for the Nash equilibria of the Asymmetric case I.*

Conclusion 2. *In the unique Nash equilibrium of the Asymmetric case II, one participant believes in winning with certainty whereas the other believes in losing with certainty. The winner believer exerts more effort under competition than under random pay. The loser believer exerts less effort under competition than under random pay.*

3.5.3. Strategic reasoning vs. lab experiment results

Our normative analysis accommodates an explanation based on strategic reasoning for the behavior of all subgroups in the lab experiment, except men under stereotype threat.

Assume male and female participants of the lab experiment face the competition strategically. If we assume that men and women believe in men's advantage to perform the multiplication task, but they believe the difference is not considerable (i.e., they both nonetheless believe that women still have a chance to

to win the competition), we are in a strategic context like Asymmetric case I. This means we can use Conclusion 1 to rationalize as strategic behavior the higher effort provision in the competitive second stage of the lab experiment of the following subgroups: i. Both male and female participants in the Implicit ST against women condition; ii. Both male and female participants in the Explicit ST against women condition; iii. Female participants in the Explicit ST against men condition.²⁸

Regarding male participants in the Explicit ST against men condition, if we take the experimental finding that men's performance improves along the second stage (from a within-subject perspective) as evidence of higher effort in the competitive stage, it could seem at a first glance that we again can rationalize men's behavior as strategic based on a context like Asymmetric case I. However, in our strategic setting we cannot rationalize the fact that men's accuracy rate is significantly lower in the competitive stage of the experiment (recall that we assume a constant (and equal) conversion rate of effort into correct answers in both stages).

Alternatively, if we take the experimental finding that men, relatively to the first stage, make virtually the same number of attempts in the competitive stage as evidence of no extra effort in the second stage, we reach the following conclusions: i. If we consider Asymmetric case I, we cannot rationalize men's behavior in the experiment since the normative prediction is an increase in effort; ii. If we instead consider Asymmetric case II²⁹, the strategic prediction is that men exert no effort in the competitive stage. We can plausibly argue that men make virtually the same number of attempts in the competitive stage of the experiment (rather than zero attempts, which would be the normative prediction) because there exists an intrinsic level of effort. However, in our view, it is implausible to argue that men's behavioral pattern characterized by a decrease in accuracy rate while holding constant the number of attempts is evidence that men are strategically not exerting their best effort in the second stage. In other words, we cannot rationalize men's behavior in the experiment based on a context like Asymmetric case II either.

3.6. Discussion

Put together, we consider that the findings of our study support the interpretation that men under stereotype threat do provide their best effort in the competitive stage, but they cannot efficiently convert their effort into correct answers.

²⁸ This strategic rationale implies that a strategic context like Asymmetric case I is preserved after the belief updating of male and female participants in the Explicit ST against women condition as well as of female participants in the Explicit ST against men condition.

²⁹ This assumes that, after receiving the stereotyping piece of information, men in the Explicit ST against men condition update their initial belief so that they believe in a considerable advantage of women to perform the multiplication task.

We advance a possible non-strategic rationale for men to exert more effort in the competitive stage of the Explicit ST against men condition. Assume that men do not reason strategically about their female opponent's mathematical ability while choosing the effort to exert in the competition. Men simply consider that the odds to win the competition are better if they exert more effort in the competitive stage. In addition, assume that earning money makes men better off, and the cost to exert effort is negligible. Then, it is rational for men to exert more effort in the competitive stage regardless of their belief about women's mathematical ability. In fact, the same non-strategic rationale can explain the extra effort in the competitive stage of any of the other subgroups.

Nevertheless, more effort is a necessary, but not a sufficient condition to improve performance in the competitive stage. A possible explanation for men's lower accuracy rate in the competitive stage when under stereotype threat is distracting thoughts. Distracting thoughts have been shown to interfere with working memory and attention, which are fundamental for the optimal performance of the multiplication task (e.g., Brewin and Smart 2005). According to the latter aspect, the cause for men's suboptimal performance in the competition when under stereotype threat is not lack of effort, but rather the fact that men unexpectedly read information that contradicts a stereotype they hold just before the start of the competition. Regarding the other subgroups, distracting thoughts do not affect men's performance in the Implicit and Explicit ST against women conditions and women in the Explicit ST against men condition because they are not the target of the threat. With respect to women in the Implicit and Explicit ST against women conditions, they do not get distracting thoughts either because the stereotype-based expectations they read do not contradict the prior belief women hold about the invoked stereotype.

3.7. Concluding remarks

In this chapter, we conduct a controlled laboratory experiment to test the hypothesis that stereotype threat affects men's and women's performance in a mixed-sex competition.

Regarding women's behavior, our results do not corroborate the psychology research on stereotype threat. According to this literature, we should expect that the gap in performance between individuals of a group prone to stereotype threat and the individuals of a group not prone to stereotype threat should be different in a situation where the threat exists compared to a situation where there is no threat. In contrast, within the context of a competition, we find that the gender gap in performance when we induce stereotype threats against women is the same as the gender gap in performance when we do not induce a stereotype threat. Nonetheless, we find evidence for the adverse effect of stereotype threat on men's competitive performance. Again within the context of a competition, we find that only women react positively to the competitive incentives when we explicitly induce stereotype-based expectations against men.

Our results regarding women are not in accordance with a stereotype threat based explanation perhaps because we evaluate the impact of mathematical stereotype threat against women in a competitive situation, whereas the psychology literature has been studying mathematical stereotype threats against women in non-competitive situations. Another possible reason for not observing the adverse effect of stereotype threat over women's performance in our study is that we use monetary incentives, which might wipe out the negative effect of stereotype threat identified in psychology research.³⁰

Our findings contradict previous results in experimental economics suggesting that men perform better under competition than women (e.g., Gneezy et al. 2003). In the three competitive conditions that we study, we observe that in two conditions men and women react similarly to competitive incentives, and in a third condition women react stronger than men to competitive incentives.

One possible explanation for this contrast in results relates to our within-subject measure of performance. In our experiment, we observe a large inter-individual heterogeneity in performance of the real-effort task. This supports our view that performance comparisons without controlling for the ability to perform the task, as is the case in a between-subject design, could lead to flawed conclusions. Another possible explanation for the contrast in results is risk attitudes. Considering the well-documented differences between men's and women's risk attitude (e.g., Croson and Gneezy 2009), it is likely that the influence of this element is higher in previous studies compared to ours because a risk-free piece-rate payment scheme has been the routinely used incentive-scheme to elicit non-competitive performance in previous studies.

An alternative interpretation of our results concerns the different pieces of information that we use in the two conditions in which we explicitly induce stereotype threat. Men and women read the same information in each of these conditions. In the Explicit ST against women condition, the information negatively stereotypes women from a women's perspective, whereas it positively stereotypes men from a men's perspective. The opposite occurs in the Explicit ST against men condition. Therefore, it is conceivable that men *excel* under competition against women when men have to *confirm* they are *better*. Men *choke* under competition against women when men have to *disconfirm* they are *worse*. Regarding women, their competitive behavior is not affected by having to disconfirm they are worse, or confirm they are better. In our view, this gender difference interpretation is less convincing because it implies that participants in the two explicit stereotype threat conditions fear the evaluation of others, which should not be the case in our experimental setting.

Our study is part of a research program that is aimed at understanding why women are underrepresented in many high-status jobs and earn lower wages than

³⁰ Using monetary incentive in a non-competitive situation, Fryer, Levitt, and List (2008) did not find evidence for the negative impact of stereotype threat either.

men. Taking into account the pervasiveness of stereotype threat, our results indicate that women do not have a weaker reaction to competitive pressure regarding performance when they *have* to compete against men. In other words, women's weaker attitude towards competition do not seem to be an explanation for the observed gender differences at the workplace if we consider the context in which men and women *are already competing* (for example, competing to be promoted).

Our results have a practical implication regarding policy design to cope with stereotype threat at the workplace. A recommendation found in the stereotype threat literature to prevent a negative effect of stereotype threat is the “stereotype nullification”, i.e., to explicitly provide individuals with information that does not conform to the stereotype (e.g., Smith and White 2002). In stark contrast, our results indicate that, at least within a competitive context, no information about the stereotype should be provided at all. If men and women are already competing against each other, they seem to cope well in terms of performance with a stereotype they hold. In this case, providing information contradicting a stereotype seems irrelevant to the ex-ante negatively stereotyped group and to harm the performance of the ex-ante positively stereotyped group.

REFERENCES

- Aronson, J., Lustina, M., Good, C., Keough, K., Steele, C., & Brown, J. (1999). When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat. *Journal of Experimental Social Psychology*, 35(1), 29–46.
- Babcock, L., & Laschever, S. (2003). *Women Don't Ask: Negotiation and the Gender*. Princeton, NJ: Princeton University Press.
- Black, S., & Strahan, P. E. (2001). The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry. *American Economic Review*, 91(4), 814–31.
- Blau, F., Ferber, M., & Winkler, A. (2013). *The Economics of Women, Men, and Work*. Englewood Cliffs, NJ: Prentice Hall.
- Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2), 542–55.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2014). Stereotypes. Working Paper.
- Brewin, C., & Smart, L. (2005). Working memory capacity and suppression of intrusive thoughts. *Journal of Behavior Therapy and Experimental Psychiatry*, 36(1), 61–68.
- Cason, T., Masters, W., & Sheremeta, R. (2010). Entry into Winner-Take-All and Proportional-Prize Contests: An Experimental Study. *Journal of Public Economics*, 94(9–10), 604–11.
- Christie, R., & Geis, F. (1970). *Scale construction*. In: *Studies in Machiavellianism*. New York: Academic Press.
- Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2), 1–27.
- Dechnaux, E., Kovenock, D., & Sheremeta, R. (2014). A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments. *Experimental Economics*, forthcoming.
- Dohmen, T., & Falk, A. (2011). Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender. *American Economic Review*, 101(2), 556–90.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. (2011). Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences. *Journal of the European Economic Association*, 9(3), 522–50.
- European Commission. (2016). *She Figures 2015*. Luxembourg: Publication Office of the European Union.
- Fischbacher, U. (2007). z-Tree Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10(2), 171–78.
- Fryer, R., Levitt, S., & List, J. (2008). Exploring the Impact of Financial Incentives on Stereotype Threat: Evidence from a Pilot Study. *American Economic Review*, 98(2), 370–75.

- Gneezy, U., Leonard, K., & List, J. (2009). Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica* , 77(5), 1637-64.
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics* , 118(3), 1049-74.
- Gneezy, U., & Rustichini, A. (2004). Gender and Competition at a Young Age. *American Economic Review Papers and Proceedings* , 94(2), 377-81.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-25.
- Günther, C., Ekinci, N., Schwier, C., & Strobel, M. (2010). Women Can't Jump? – An Experiment on Competitive Attitudes and Stereotype Threat. *Journal of Economic Behavior & Organization* , 75 (3), 395-401.
- Gupta, N., Poulsen, A., & Villeval, M. (2013). Gender matching and competitiveness: experimental evidence. *Economic Inquiry* , 51(1), 816-35.
- Hillman, A., & Riley, J. (1989). Politically Contestable Rents and Transfers. *Economics and Politics* , 1(1), 17-39.
- Hyde, J., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin* , 107(2), 139-55.
- Inzlicht, M., & Schmader, T. (2012). *Stereotype threat: Theory, process, and application*. New York: Oxford Univ. Press.
- Iriberry, N., & Rey-Biel, P. (2017). Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *Journal of Economic Behavior & Organization*, 135, 99-111.
- Konrad, K. (2009). *Strategy and Dynamics in Contests*. Oxford University Press.
- Lazear, E., & Rosen, S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy* , 89(5), 841-64.
- Moldovanu, B., & Sela, A. (2001). The Optimal Allocation of Prizes in Contests. *American Economic Review* , 91(3), 542-58.
- Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics* , 122(3), 1067-1101.
- Offerman, T. (1997). *Beliefs and Decision Rules in Public Good Games*. Dordrecht: Kluwer Academic Publishers.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology* , 85(3), 440-52.
- Smith, J., & White, P. (2002). An Examination of implicitly activated, explicitly activated and nullified stereotypes on mathematical performance: It's not just a female's issue. *Sex Roles: A Journal of Research* , 47(3-4), 179–91.
- Spencer, S., Logel, C., & Davies, P. (2016). Stereotype threat. *Annual review of psychology*, 67, 415-37.

- Spencer, S., Steele, C., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28.
- Steele, C., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Tullock, G. (1980). Efficient Rent-Seeking. In J. Buchanan, R. Tollison, & G. Tullock, *Toward a theory of the rent seeking-society*. College Station: Texas A & M University.

Appendix 3.A.1.

Additional notes on the experimental procedure of the lab experiment

3.A.1.1. The quadratic scoring rule incentive scheme

Participants in the lab experiment read the following instructions right after the non-competitive first stage performance:

“Instructions: Rank your performance in the previous stage

Four other participants have been randomly chosen.

*Now you have to indicate your best estimates that exactly 0, exactly 1, exactly 2, exactly 3 or exactly 4 of these four participants correctly solved **more** problems than you did in the previous 5 minutes performance stage.*

You have to indicate your best estimates in percentages by typing an integer number between 0 and 100 (inclusive). The sum of the numbers you type in has to add up to 100.

Example 1: *If you are completely sure that exactly two of the four randomly chosen participants were better than you, then you have to type the number 0 in the box of "Exactly zero" and in the box of "Exactly one", type the number 100 in the box of "Exactly two", and type the number 0 in the two remaining boxes.*

Example 2: *If you are completely unsure about all the options, then you have to type the number 20 in all boxes. This would mean you estimate that all options are equally likely.*

Example 3: *If you are neither completely sure nor completely unsure about your performance relative to the performance of the four randomly chosen participants you can indicate this by choosing other percentages. For instance, if you estimate that there is a 10 percent chance that exactly 0, a 15 percent chance that exactly 1, a 34 percent chance that exactly 2, a 37 percent chance that exactly 3, and a 4 percent chance that exactly 4 of the other four participants correctly solved more problems, you can indicate this by typing in these numbers in the corresponding boxes.*

The numbers in the examples are just for illustrative purposes. They DO NOT intend to suggest how anyone may choose among the different options.

You will be paid for the accuracy of your estimates. *These earnings depend on the estimates you indicate and on the true outcome. The earnings are calculated according to a mathematical formula. It is not important for you to understand this formula now, but you can inspect it when the experiment is over. What is important for you to know is that your earnings are maximized if you indicate your prediction truthfully. If your estimates are completely correct, you earn € 3. If your estimates are completely wrong, you earn € 0. If your estimates are neither completely correct nor completely wrong, you earn something between € 0 and € 3. You earn more the closer your estimates are to the true outcome.*

At the end of the experiment you will be informed about these earnings. If you have a question, please raise your hand."

At the end of the experiment, upon a participant's request, we not only show the formula but also how the formula works. The formula that we use to determine each participant's earnings is:

$$1.5 + 3 \times \frac{E_{actual\ rank}}{100} - 1.5 \times \left[\left(\frac{E_{rank\ 1}}{100} \right)^2 + \left(\frac{E_{rank\ 2}}{100} \right)^2 + \left(\frac{E_{rank\ 3}}{100} \right)^2 + \left(\frac{E_{rank\ 4}}{100} \right)^2 + \left(\frac{E_{rank\ 5}}{100} \right)^2 \right]$$

In which:

$E_{rank\ 1}$ represents the estimate that exactly 0 of the other four participants correctly solved more problems

$E_{rank\ 2}$ represents the estimate that exactly 1 of the other four participants correctly solved more problems

$E_{rank\ 3}$ represents the estimate that exactly 2 of the other four participants correctly solved more problems

$E_{rank\ 4}$ represents the estimate that exactly 3 of the other four participants correctly solved more problems

$E_{rank\ 5}$ represents the estimate that exactly 4 of the other four participants correctly solved more problems

$E_{actual\ rank}$ represents the rank estimate corresponding to the participant's actual rank

Offerman (1997) shows that this mechanism is incentive compatible to elicit truthful beliefs.

3.A.1.2. Mach IV test

The Mach IV test is a personality survey that consists of the following twenty-statements:

1. *Never tell anyone the real reason you did something unless it is useful to do so.*
2. *The best way to handle people is to tell them what they want to hear*
3. *One should take action only when one is sure it is morally right.*
4. *Most people are basically good and kind.*
5. *It is safest to assume that all people have a malicious personality and it will come out when they are given a chance.*
6. *Honesty is the best policy in all cases.*
7. *There is no excuse for lying to someone else.*
8. *Generally speaking, people won't work hard unless they're forced to do so.*
9. *All in all, it is better to be humble and honest than to be important and dishonest.*
10. *When you ask someone to do something for you, it is best to give the real reasons for wanting it rather than giving reasons which carry more weight.*
11. *Most of the successful people in the world have clean, moral lives.*
12. *Anyone who completely trusts anyone else is asking for trouble.*
13. *The biggest difference between most criminals and other people is that the criminals are stupid enough to get caught.*
14. *Most people are brave.*
15. *It is wise to flatter important people.*
16. *It is possible to be good in all respects.*
17. *To say that there is a sucker born every minute is wrong.*
18. *It is hard to be successful without cutting corners here and there.*
19. *People suffering from incurable diseases should have the choice of being put painlessly to death.*
20. *Most people forget more easily the death of their parents than the loss of their property.*

For each statement, we ask participants to “*please indicate the degree to which you personally agree or disagree*” using the following 7-point Likert scale:

“*strongly disagree*” | “*disagree*” | “*slightly disagree*” | “*neutral*” | “*slightly agree*” | “*agree*” | “*strongly agree*”

For statements 1, 2, 5, 8, 12, 13, 15, 18, 19 and 20, the score is determined as follows:

“*strongly disagree*” (**1 point**) | ... | “*strongly agree*” (**7 points**)

For statements 3, 4, 6, 7, 9, 10, 11, 14, 16 and 17, the score is determined as follows:

“*strongly disagree*” (**7 points**) | ... | “*strongly agree*” (**1 point**)

Thus, the test score ranges between 20-140 points.

Appendix 3.A.2. Design of the online questionnaire

We recruit the prospective respondents to the online questionnaire from the same subject pool of the participants in the lab experiment. We separately invite men and women to participate in the online questionnaire, and we set a participation quota of 60 respondents for each sex. More specifically, registration to respond to the online questionnaire automatically ends once 60 male (female) respondents complete the questionnaire. In addition, we inform in the invitations that the questionnaire takes around 15-minutes and that respondents can earn money with their answers.

The questionnaire consists of two parts. In the first part of the questionnaire, respondents first read the instructions of the lab experiment that explain the multiplication task and the 5 minutes performance under a random pay incentive scheme. To ensure the understanding of respondents, we ask a series of control questions about the multiplication task and the random pay incentive scheme. Respondents could only proceed with the questionnaire if they correctly answer the control questions.

Subsequently, we ask respondents to make the following (monetarily incentivized) estimates regarding the behavior of participants in the lab experiment:

“For your information: In the BEElab-experiment the minimum number of multiplication problems correctly solved during the 5 minutes by any participant was 3 and the maximum number was 57.

With your answer to the following question you can earn money. Please provide your best estimate on the following.

How many problems did MALE participants in the described BEElab-experiment correctly solve, on average, during the 5 minutes?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual average number of problems (rounded to integer numbers) male participants correctly solved in the BEElab-experiment. If your estimate is equal to the actual average number, you will earn € 40. If your estimate deviates by plus or minus 1 from the actual average number, you will earn € 20. If your estimate deviates by plus or minus 2 from the actual average number, you will earn € 10. If your estimate deviates by more than plus or minus 2 from the actual average number, you will earn nothing.”

“For your information: In the BEElab-experiment the minimum number of multiplication problems correctly solved during the 5 minutes by any participant was 3 and the maximum number was 57.

With your answer to the following question you can earn money. Please provide your best estimate on the following.

How many problems did FEMALE participants in the described BEElab-experiment correctly solve, on average, during the 5 minutes?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual average number of problems (rounded to integer numbers) female participants correctly solved in the BEElab-experiment. If your estimate is equal to the actual average number, you will earn € 40. If your estimate deviates by plus or minus 1 from the actual average number, you will earn € 20. If your estimate deviates by plus or minus 2 from the actual average number, you will earn € 10. If your estimate deviates by more than plus or minus 2 from the actual average number, you will earn nothing.”

We present the questions in separate screens. Half of the male respondents (30) first estimate the performance of the male participants. Then, in a subsequent screen, they estimate the performance of the female participants. Vice-versa for the other half of the male respondents. We use the same procedure with respect to the female respondents. Since we did not find any order effect, we report the results of the questionnaire only with respect to the pooled data.

The second part of the questionnaire splits into two versions. In one version, 30 male and 30 female respondents unexpectedly see during 40s the piece of information that we used in the lab experiment to induce stereotype-based expectations against women. In the other version, the remaining 30 male and 30 female respondents unexpectedly see during 40s the piece of information that we used in the lab experiment to induce stereotype-based expectations against men.

Subsequently, in both versions, we ask respondents whether they want to change their previous estimates. Specifically, respondents read the following instructions:

“If you wish you now can change your previous estimate(s).

You will be paid either according to your previous estimates if you do not make any change, or according to your new estimate(s) if you change your previous estimate(s).

Your previous estimates were:

Male participants: # problems correctly solved

Female participants: # problems correctly solved

*Would you like to change your previous estimate for **men**? Yes No*

*Would you like to change your previous estimate for **women**? Yes No”*

If the respondent answers “No” to both questions, the questionnaire ends. If the respondent answers “Yes” to change one (or both) of the previous estimates, the corresponding question(s) of part one of the questionnaire appears again in the subsequent screen.

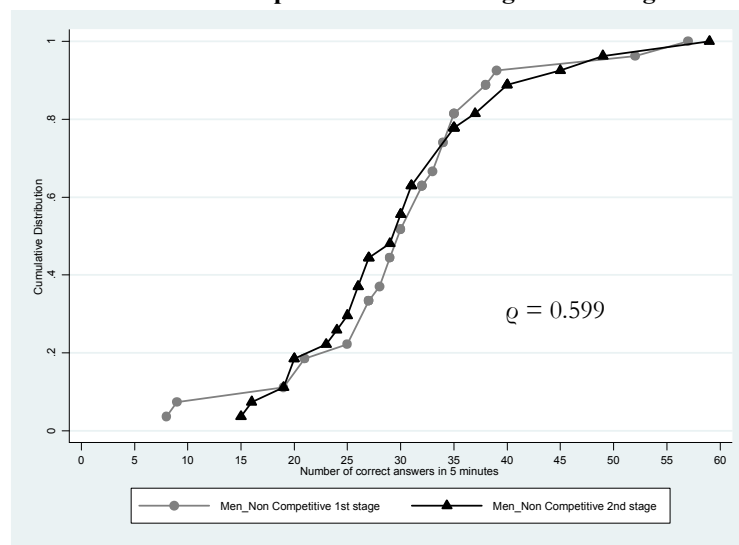
We conduct the online questionnaire via Qualtrics Online Survey software. The respondents are predominantly (83%) students of Business and Economics at Maastricht University. Average earnings are € 17.

Appendix 3.A.3. Additional data of the lab experiment

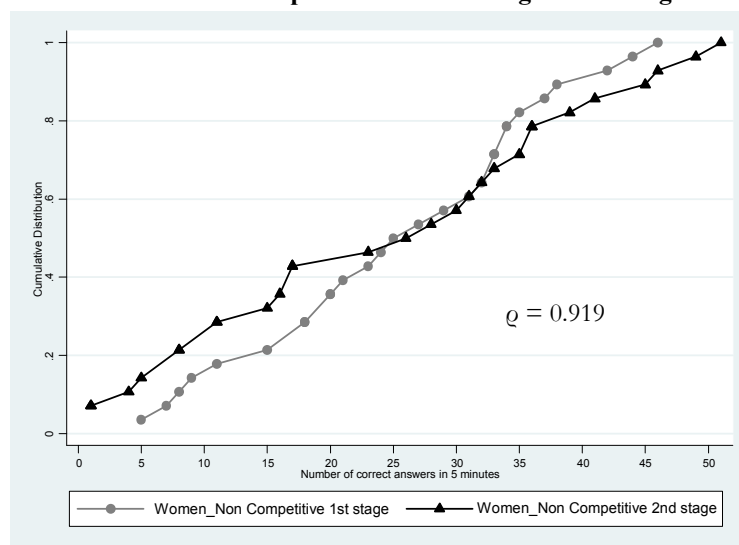
3.A.3.1. Cumulative distributions (CDs) of performance in the Twice random pay condition

Note: The 1st stage performance is under non-competitive incentives. The 2nd stage performance is under competitive incentives. ρ indicates the Spearman rank correlation of performance between stages.

i. CDs of men's performance: 1st stage vs. 2nd stage



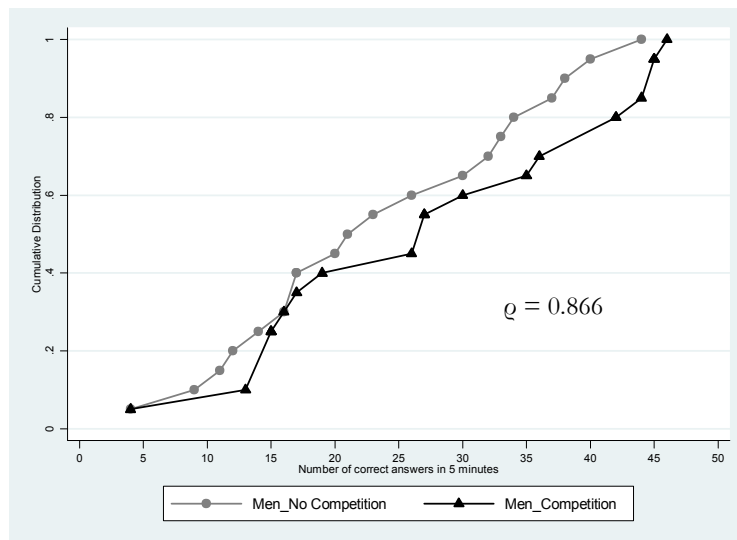
i. CDs of women's performance: 1st stage vs. 2nd stage



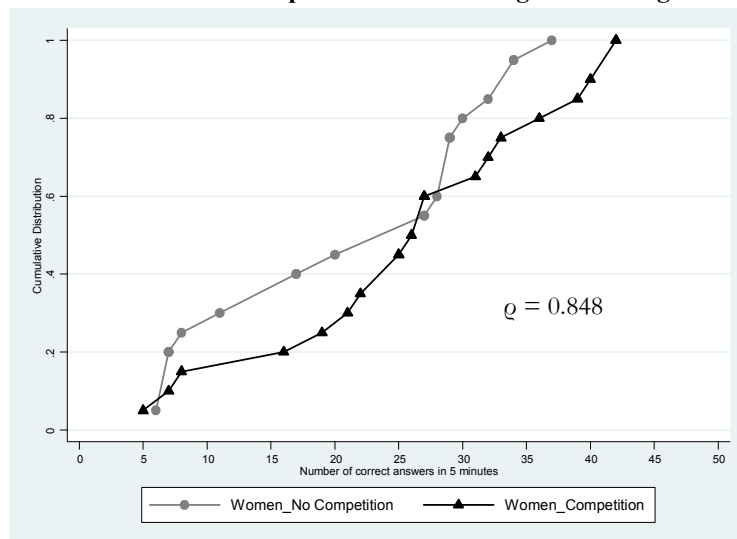
3.A.3.2. Cumulative distributions (CDs) of performance in the Implicit ST against women condition

Note: The 1st stage performance is under non-competitive incentives. The 2nd stage performance is under competitive incentives. ρ indicates the Spearman rank correlation of performance between stages.

i. CDs of men's performance: 1st stage vs. 2nd stage



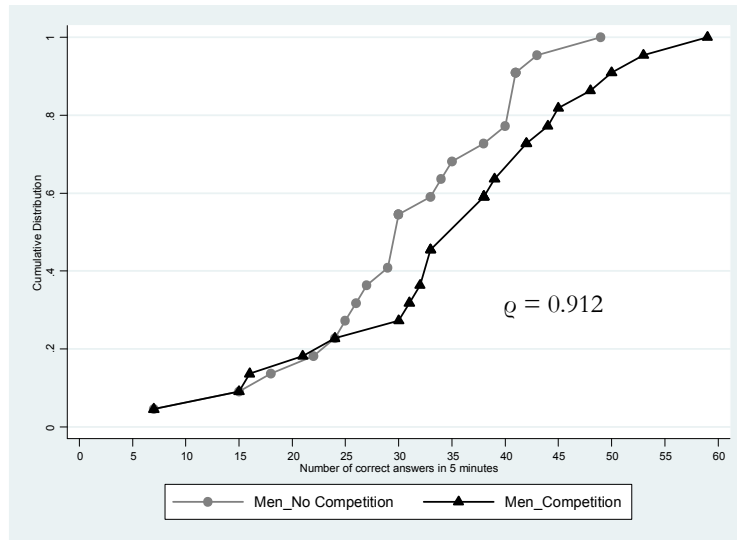
i. CDs of women's performance: 1st stage vs. 2nd stage



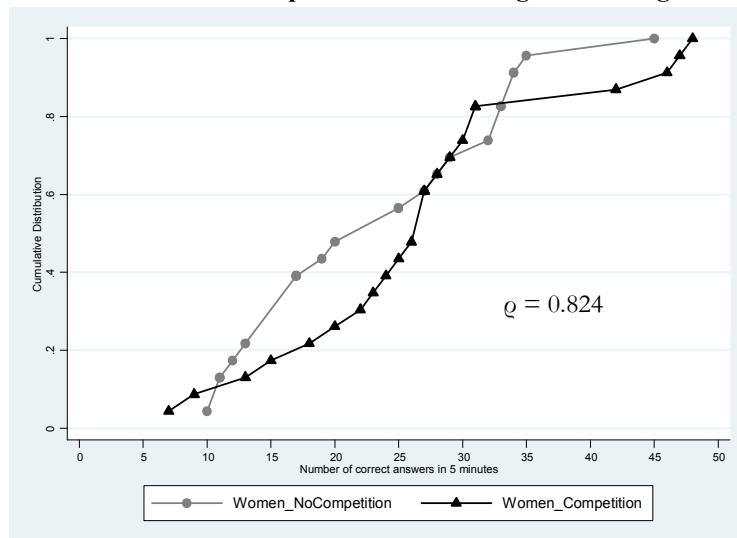
3.A.3.3. Cumulative distributions (CDs) of performance in the Explicit ST against women condition

Note: The 1st stage performance is under non-competitive incentives. The 2nd stage performance is under competitive incentives. ρ indicates the Spearman rank correlation of performance between stages.

i. CDs of men's performance: 1st stage vs. 2nd stage



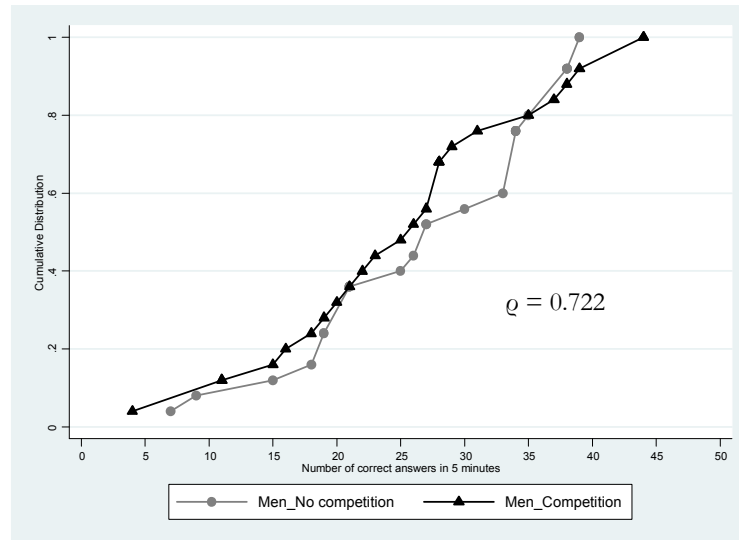
i. CDs of women's performance: 1st stage vs. 2nd stage



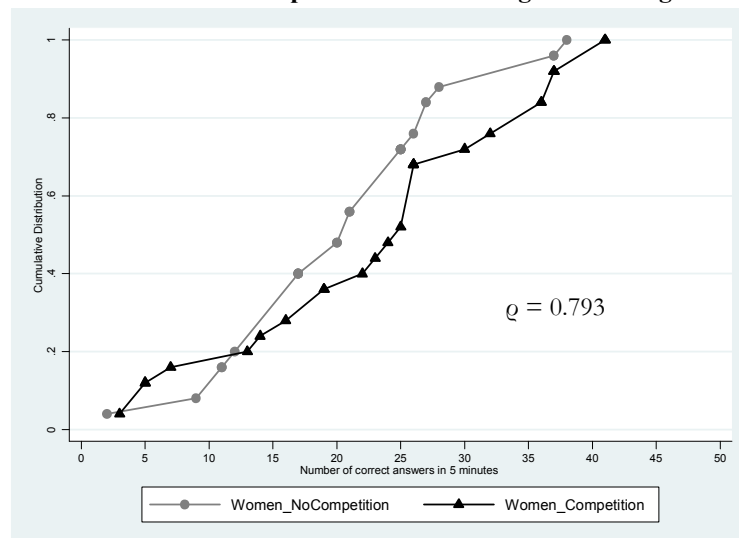
3.A.3.4. Cumulative distributions (CDs) of performance in the Explicit ST against men condition

Note: The 1st stage performance is under non-competitive incentives. The 2nd stage performance is under competitive incentives. ρ indicates the Spearman rank correlation of performance between stages.

i. CDs of men's performance: 1st stage vs. 2nd stage



i. CDs of women's performance: 1st stage vs. 2nd stage



Appendix 3.A.4.

Effort provision—A normative view | Detailed analysis

Consider an arbitrary participant i of the lab experiment. We assume that participant i 's number of correct answers in each stage equals e_i , which represents the effort participant i exerts. In line with the lab experiment, the size of the monetary prize participant i may earn in each performance stage is not fixed ex-ante, but rather depends positively on effort provision. Specifically, participant i 's revenue is represented by $R_i(e_i) = 0.40 \cdot e_i$ in which 0.40 indicates participant i 's earnings per correct answer in Euro cents, $R'_i(e_i) > 0$ and $R''_i(e_i) = 0$. Therefore, we assume that the accuracy rate to convert effort into correct answers is the same in both stages. Moreover, participant i 's cost to exert effort is represented by a strictly convex function $c_i : R_0^+ \rightarrow R_0^+$. Specifically, we assume that $c_i(e_i) = \frac{\alpha_i}{2} \cdot e_i^2$ in which $\alpha_i > 0$ indicates participant i 's ability to correctly solve multiplication problems, $c_i(0) = 0$, $c'_i(e_i) > 0$ for $e_i > 0$ and $c''_i(e_i) > 0$.

3.A.4.1. First stage: performance under a random pay incentive scheme

Participant i faces the following problem in the first stage of the lab experiment:

$$(1) \quad \max_{e_i} \pi_i(e_i) \quad \text{with} \quad \pi_i(e_i) = \text{prob}^{rp} \cdot 0.4 \cdot e_i - \frac{\alpha_i}{2} \cdot e_i^2$$

in which $\text{prob}^{rp} = 0.5$ represents the probability of being paid under the random pay incentive scheme.

Observation 1. *Participant i 's optimal level of effort under the random pay incentive scheme equals $e_i^{rp*} = \frac{0.2}{\alpha_i}$.*

Proof. The f.o.c. of problem (1) is:

$$\pi'_i(e_i) = 0 \Leftrightarrow 0.5 \cdot 0.4 - \alpha_i \cdot e_i = 0 \Leftrightarrow e_i^{rp*} = \frac{0.2}{\alpha_i}$$

in which e_i^{rp*} is the participant i 's optimal level of effort under the random pay incentive scheme. QED

Not surprisingly, participant i 's optimal effort level under random pay is increasing in participant i 's ability to correctly answer problems (therefore decreasing in α_i).

3.A.4.2. Second stage: performance under a winner-take-all tournament incentive scheme

The second stage of the lab experiment consists of a one-to-one mixed-sex competition in which participants perform the multiplication task during 5 minutes under a winner-take-all tournament incentive scheme. To analyze the strategic nature of this stage, we consider a static game of complete information.¹ Specifically, we consider a static game in which both players, to whom we generically refer to as participant i and participant j , have to simultaneously choose their effort level before the start of the competition. Thus, a possible outcome of the game is any pair of strategies (e_i^{com}, e_j^{com}) in which e_i^{com} and e_j^{com} are the effort levels that participant i and participant j , respectively, choose in the competitive second stage. Without loss of generality, we proceed the exposition considering participant i 's perspective.

In this competitive setting, participant i probability of winning the tournament against participant j is described by the following *Tournament Success Function*:

$$prob_i^{com}(e_i^{com}, e_j^{com}) = \begin{cases} 1 & \text{if } e_i^{com} \geq e_j^{com} \\ 0 & \text{if } e_i^{com} < e_j^{com} \end{cases}$$

in which $prob_i^{com}$ represents the probability of participant i winning the competition against participant j , i.e., the probability of being paid under the winner-take-all tournament incentive scheme.²

Hence, participant i faces the following problem in the second stage of the lab experiment:

$$(2) \quad \max_{e_i} \pi_i(e_i | b_{i, e_j^{com}}) \quad \text{with} \quad \pi_i(e_i | b_{i, e_j^{com}}) = prob_i^{com}(e_i | b_{i, e_j^{com}}) \cdot 0.4 \cdot e_i - \frac{\alpha_i}{2} \cdot e_i^2$$

¹ It would certainly be more realistic to consider a static game of incomplete information, in which participants are privately informed about their ability parameter a . In section 3.A.7., we concisely depict the economics literature on contests and tournaments to explain why we instead put forth a static game of complete information.

² Strictly speaking, the participants in the lab experiment equally split the total earnings in case of a tie. In our analysis, we instead assume that both participants still get the total earnings in case of a tie. Nevertheless, we report the consequences for the Nash equilibria that we find if we did not make this simplifying assumption.

in which $b_{i,e_j^{com}}$ represents participant i 's belief about the effort participant j will exert in the competitive stage. Thus, $prob_i^{com}(e_i | b_{i,e_j^{com}}) = 1$ if $e_i^{com} \geq b_{i,e_j^{com}}$ or $prob_i^{com}(e_i | b_{i,e_j^{com}}) = 0$ if $e_i^{com} < b_{i,e_j^{com}}$.

Differently to the first stage, participant i now takes into consideration the behavior of participant j when choosing the effort level to exert in the competition. Specifically, participant i holds beliefs about e_j , which depend on α_i and α_j , the participant i 's and participant j 's ability to correctly answer problems, respectively.

To determine Nash equilibria in our competitive setting, we separately analyze three cases that are collectively exhaustive to derive the results: i. The *Symmetric case*, in which participant i 's ability equals participant j 's ($\alpha_i = \alpha_j$); ii. The *Asymmetric case I*, in which participant i 's ability is higher than participant j 's ($\alpha_i < \alpha_j$), but the difference in ability is not considerable (in the sense that both participants nonetheless believe that participant j still has a chance to win the competition); iii. The *Asymmetric case II*, in which participant i 's ability is considerably higher than participant j 's ($\alpha_i \ll \alpha_j$).

Before analyzing each case, we first claim:

Observation 2. *The maximum level of effort participant i could rationally exert during the competitive second stage is the effort level at which the total revenue equals total cost. Formally, $e_i^{com} \leq \hat{e}_i$ for any $b_{i,e_j^{com}}$ in which $\hat{e}_i = \arg \left[R_i(e_i) = c_i(e_i) \right]$. In other words, any strategy $e_i^{com} > \hat{e}_i$ cannot be part of a Nash equilibrium.*

Proof. Consider that participant i exerts $e_i^{com} > \hat{e}_i$. Then,

1. $\pi_i < 0$ regardless participant i wins or loses the competition because $R_i(e_i) < c_i(e_i)$ for $e_i^{com} > \hat{e}_i$, i.e., participant i 's expected total revenue is lower than expected total cost.
2. $\pi_i = 0$ if instead participant i exerts $e_i^{com} = 0$.

Therefore, any strategy $e_i^{com} > \hat{e}_i$ cannot be part of a Nash equilibrium because it is strictly dominated by $e_i^{com} = 0$. QED

Therefore, we define $E_i = [0, \hat{e}_i]$ as the rational actions/strategies space of participant i in this game. Note that the higher is participant i 's ability (which

corresponds to a lower α_i and, as a result, a higher \hat{e}_i), the wider is participant i 's rational actions/strategies space.

3.A.4.2.1. Symmetric case ($\alpha_i = \alpha_j$)

In this case, it is common knowledge for both participants that $c_i(e_i) = c_j(e_j)$, which implies that $\hat{e}_i = \hat{e}_j$ (see section 3.A.5.1 for a graphical description). For a given belief about e_j , participant i compares the expected payoffs of losing or winning the tournament in order to decide the best response level of effort. Thus, we subsequently consider the loser scenario and the winner scenario.

▪ **Loser scenario.** If participant i exerts an effort level lower than their belief about participant j 's effort (i.e., $e_i^{com} < b_{i,e_j^{com}}$), then participant i believes in losing the tournament with certainty. Thus, in this scenario participant i 's problem (2) becomes:

$$\max_{e_i} -\frac{\alpha_i}{2} \cdot e_i^2$$

The trivial solution is $e_i^{com^L} = 0$, in which $e_i^{com^L}$ represents participant i 's optimal level of effort when a loser. In short, when participant i believes to be the loser, the optimal response is to exert zero effort, which yields a payoff of zero.

▪ **Winner scenario.** If participant i exerts an effort level higher or equal to their belief about participant j 's effort (i.e., $e_i^{com} \geq b_{i,e_j^{com}}$), then participant i believes in winning the tournament with certainty. Thus, in this scenario participant i 's problem (2) becomes:

$$\max_{e_i} 0.4 \cdot e_i - \frac{\alpha_i}{2} \cdot e_i^2$$

The solution of the problem is (see derivation of (3) in section 3.A.5.1):

$$(3) \quad e_i^{com^W} = \begin{cases} e_i^{\text{tangent}} & \text{if } 0 \leq b_{i,e_j^{com}} \leq e_i^{\text{tangent}} \\ b_{i,e_j^{com}} & \text{if } e_i^{\text{tangent}} < b_{i,e_j^{com}} \leq \hat{e}_i \\ b_{i,e_j^{com}} & \text{if } b_{i,e_j^{com}} > \hat{e}_i \end{cases}$$

in which $e_i^{com^W}$ represents participant i 's optimal level of effort when a winner and

$e_i^{\text{tangent}} = \frac{0.4}{\alpha_i}$ is the effort level at which the expected marginal revenue and marginal cost of correctly answering problems are equal.

Combining the results of the two scenarios, we obtain participant i 's best response function (see derivation of (4) in section 3.A.5.1):

$$(4) \quad BR_i(b_{i,e_j^{com}}) = \begin{cases} e_i^{com*} = e_i^{\text{tangent}} & \text{if } 0 \leq b_{i,e_j^{com}} \leq e_i^{\text{tangent}} \\ e_i^{com*} = b_{i,e_j^{com}} & \text{if } e_i^{\text{tangent}} < b_{i,e_j^{com}} \leq \hat{e}_i \\ e_i^{com*} = 0 & \text{if } b_{i,e_j^{com}} > \hat{e}_i \end{cases}$$

in which e_i^{com*} represents participant i 's optimal level of effort in the competitive second stage.

Participant j 's problem is symmetric. Therefore, participant j 's best response function is:

$$BR_j(b_{j,e_i^{com}}) = \begin{cases} e_j^{com*} = e_j^{\text{tangent}} & \text{if } 0 \leq b_{j,e_i^{com}} \leq e_j^{\text{tangent}} \\ e_j^{com*} = b_{j,e_i^{com}} & \text{if } e_j^{\text{tangent}} < b_{j,e_i^{com}} \leq \hat{e}_j \\ e_j^{com*} = 0 & \text{if } b_{j,e_i^{com}} > \hat{e}_j \end{cases}$$

in which e_j^{com*} represents participant j 's optimal level of effort in the competitive second stage and $e_j^{\text{tangent}} = \frac{0.4}{\alpha_j}$.

Based on the participant i 's and j 's best response function, we state our first proposition:

Proposition 1. *Any pair of strategies $(e_i^{com}, e_j^{com}) \in [e_i^{\text{tangent}}, \hat{e}_i]$ such that $e_i^{com} = e_j^{com}$ is a Nash equilibrium of the Symmetric case.³*

³ Evidently, this is equivalent to state that $\forall (e_i^{com}, e_j^{com}) \in [e_j^{\text{tangent}}, \hat{e}_j]$ such that $e_i^{com} = e_j^{com}$ is a Nash equilibrium of the Symmetric case. Moreover, if we instead assumed that the participants equally split

Proof. See proof in section 3.A.6.

3.A.4.2.2. Asymmetric case I ($\alpha_i < \alpha_j$ but $\hat{e}_j \geq e_i^{\text{tangent}}$)⁴

In this case, it is common knowledge for both participants that participant i has higher ability (i.e., $c_i(e_i) < c_j(e_j)$), which implies that $\hat{e}_i > \hat{e}_j$ (see section 3.A.5.2 for a graphical description). Nevertheless, it is also common knowledge that $\hat{e}_j \geq e_i^{\text{tangent}}$, which means that the difference in ability is not *considerable*.

The analytical representation of the best response function of participant i and participant j , respectively, is the same as in the Symmetric case. Yet, since now $e_i^{\text{tangent}} > e_j^{\text{tangent}}$ and $\hat{e}_i > \hat{e}_j$ the incentives to deviate are different (see section 3.A.5.2 for the analysis of profitable deviations in the Asymmetric case I). Our second proposition is:

Proposition 2. *Any pair of strategies $(e_i^{\text{com}}, e_j^{\text{com}}) \in [e_i^{\text{tangent}}, \hat{e}_j]$ such that $e_i^{\text{com}} = e_j^{\text{com}}$ is a Nash equilibrium of the Asymmetric case I.*⁵

Proof. See proof in section 3.A.6.

3.A.4.2.3. Asymmetric case II ($\alpha_i < \alpha_j \Leftrightarrow \alpha_i < \alpha_j$ and $\hat{e}_j < e_i^{\text{tangent}}$)

In this case, it is common knowledge for both participants not only that participant i has higher ability (i.e., $c_i(e_i) < c_j(e_j)$) but also that $\hat{e}_j < e_i^{\text{tangent}}$, which means that the difference in ability is *considerable* (see section 3.A.5.3 for a graphical description).

Hence, it is common knowledge for both participants that the maximum effort level participant j could rationally exert is within the range of effort levels where participant i 's marginal payoff is still increasing. Thus, at least one of the participants has an incentive to deviate, but for the pair of strategies that we state in our third proposition:

Proposition 3. $(e_i^{\text{com}}, e_j^{\text{com}}) = (e_i^{\text{tangent}}, 0)$ is the unique Nash equilibrium of the Asymmetric case II.⁶

the total earnings in case of a tie, no Nash equilibrium in pure strategies would exist in the Symmetric case (proof available upon request).

⁴ Without loss of generality, we assume participant i has higher ability. The conclusions we reach in both the Asymmetric case I and the Asymmetric case II hold the other way around if we instead assumed participant j has higher ability.

⁵ If we instead assumed that the participants equally split the total earnings in case of a tie, no Nash equilibrium in pure strategies would exist in the Asymmetric case I (proof available upon request).

⁶ If we instead assumed that the participants equally split the total earnings in case of a tie, the unique Nash equilibrium in pure strategies would be exactly the same (proof available upon request).

Proof. See proof in section 3.A.6.

3.A.4.3. Effort under random pay vs. effort under competition

We draw two important observations from the sections 3.A.4.1 and 3.A.5.2:

Observation 3. *In each Nash equilibrium of the Symmetric case, the participants exert more effort under competition than under random pay. The same holds for the Nash equilibria of the Asymmetric case I.*

Proof. With reference to the set of Nash equilibria in the Symmetric case, the Nash equilibrium involving the lowest level of effort is $(e_i^{com}, e_j^{com}) = (e_i^{tangent}, e_j^{tangent})$ where $e_i^{tangent} = e_j^{tangent} = \frac{0.4}{\alpha_i} = \frac{0.4}{\alpha_j}$.

With reference to the set of Nash equilibria in the Asymmetric case I, the Nash equilibrium involving the lowest level of effort is $(e_i^{com}, e_j^{com}) = (e_i^{tangent}, e_i^{tangent})$ where $e_i^{tangent} = \frac{0.4}{\alpha_i} > e_j^{tangent} = \frac{0.4}{\alpha_j}$ because $\alpha_i < \alpha_j$.

In both cases, the optimal effort level of participant i and participant j under competition is higher than their optimal effort level under random pay since $\frac{0.4}{\alpha_i} > e_i^{rp*} = \frac{0.2}{\alpha_i}$ and $\frac{0.4}{\alpha_i} > e_j^{rp*} = \frac{0.2}{\alpha_j}$, respectively. QED

Observation 4. *In the unique Nash equilibrium of the Asymmetric case II, one participant believes in winning with certainty whereas the other believes in losing with certainty. The winner believer exerts more effort under competition than under random pay. The loser believer exerts less effort under competition than under random pay.*

Proof. Without loss of generality, consider that participant i is the participant with higher ability in Asymmetric case II.

Then, participant i is the winner believer and exerts $e_i^{com^W} = \frac{0.4}{\alpha_i}$ under competition, which is higher than $e_i^{rp*} = \frac{0.2}{\alpha_i}$, the optimal effort of participant i under random pay.

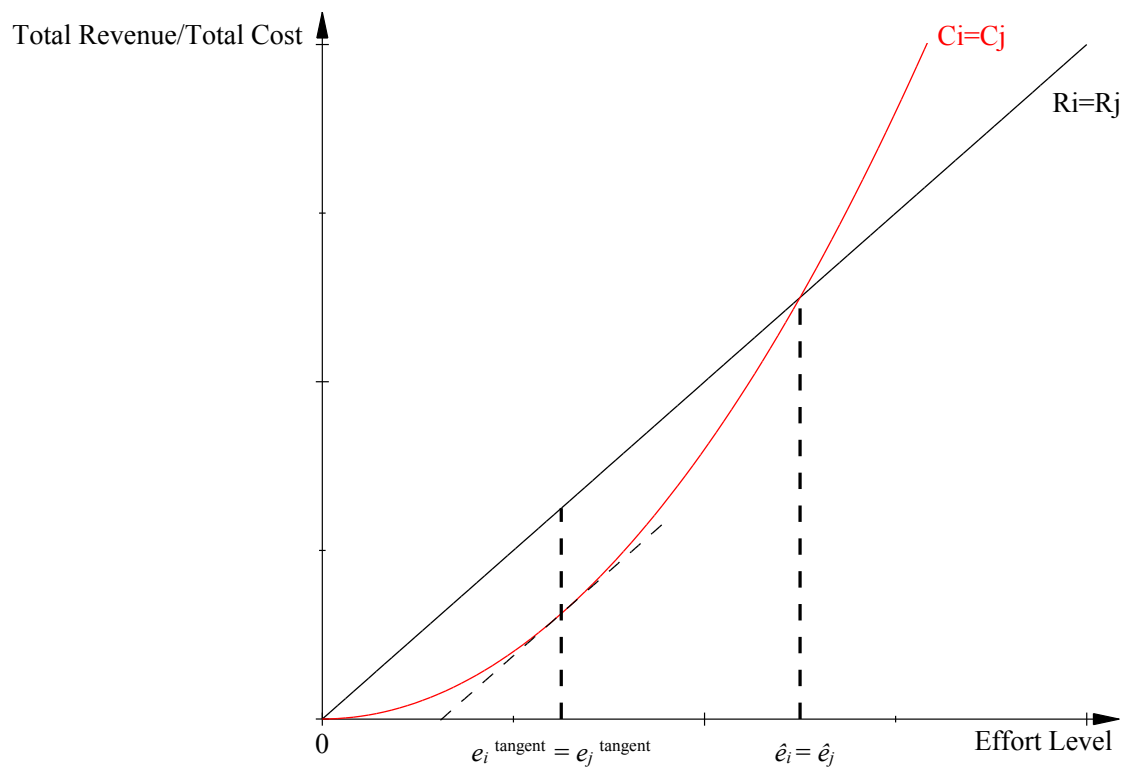
On the other hand, participant j is the loser believer and exerts $e_j^{com^L} = 0$ under competition, which is lower than $e_j^{rp*} = \frac{0.2}{\alpha_j}$, the optimal effort of participant j under random pay. QED

Appendix 3.A.5. Graphical description and derivation of best response functions

3.A.5.1. Symmetric case

In Figure 3.A.5.1, we depict qualitatively the symmetric case:

Figure 3.A.5.1. Symmetric case



We now derive participant i 's best response function. Following the exposition in section 3.A.4.2.1., we first separately derive the solution for the loser and winner scenarios.

Participant i 's problem in the loser scenario ($e_i^{com} < b_{i,e_j^{com}}$):

$$\max_{e_i} -\frac{\alpha_i}{2} \cdot e_i^2$$

The f.o.c. is:

$$\pi'_i(e_i) = 0 \Leftrightarrow -\alpha_i \cdot e_i = 0 \Leftrightarrow e_i^{com^L} = 0$$

in which $e_i^{com^L}$ represents participant i 's optimal level of effort if a loser.

Participant i 's problem in the winner scenario ($e_i^{com} \geq b_{i,e_j^{com}}$):

$$\max_{e_i} 0.4 \cdot e_i - \frac{\alpha_i}{2} \cdot e_i^2$$

The f.o.c. is:

$$(1) \quad \pi'_i(e_i) = 0 \Leftrightarrow 0.4 - \alpha_i \cdot e_i = 0 \Leftrightarrow e_i^{com^W} = \begin{cases} e_i^{\text{tangent}} = \frac{0.4}{\alpha_i} & \text{if } 0 \leq b_{i,e_j^{com}} \leq e_i^{\text{tangent}} \\ b_{i,e_j^{com}} & \text{if } e_i^{\text{tangent}} < b_{i,e_j^{com}} \leq \hat{e}_i \\ b_{i,e_j^{com}} & \text{if } b_{i,e_j^{com}} > \hat{e}_i \end{cases}$$

in which $e_i^{com^W}$ represents participant i 's optimal level of effort if a winner and $e_i^{\text{tangent}} = \frac{0.4}{\alpha_i}$ is the effort level at which the expected marginal revenue and marginal cost of correctly answering problems are equal.

Simultaneously analyzing the solutions of the two scenarios, we now derive participant i 's best response function.

Claim. For any participant i 's belief such that $0 \leq b_{i,e_j^{com}} \leq e_i^{\text{tangent}}$, participant i 's best response is to win the tournament and exert $e_i^{com^W} = e_i^{\text{tangent}} = \frac{0.4}{\alpha_i}$.

Proof. With reference to the Figure 3.A.5.1, we see that in the range $[0, e_i^{\text{tangent}})$ the marginal cost of effort is always lower than the expected marginal revenue of effort. Thus, the more participant i *overcuts* participant j in this range, participant i not only wins the tournament but also increases their payoff as the winner. Moreover, for any $b_{i,e_j^{com}} \in [0, e_i^{\text{tangent}}]$, e_i^{tangent} is always participant i 's best response because it yields a payoff higher than zero, which is participant i 's payoff in case he/she rather decided to lose the tournament. QED

However, exerting e_i^{tangent} could be not sufficient to win the competition. We describe this possibility in the second branch of (1).

Claim. In case participant i 's belief is such that $e_i^{\text{tangent}} < b_{i,e_j^{\text{com}}} \leq \hat{e}_i$, then participant i 's best response is to win the tournament and exert $e_i^{\text{com}^W} = b_{i,e_j^{\text{com}}}$

Proof. With reference again to the 3.A.5.1, we see that in the range $(e_i^{\text{tangent}}, \hat{e}_i]$ the marginal cost of effort is always higher than the expected marginal revenue of effort. Thus, any extra effort participant i exerts on top of the effort level sufficient to win the tournament leads to a decrease in their payoff as the winner. Moreover, for any $b_{i,e_j^{\text{com}}} \in (e_i^{\text{tangent}}, \hat{e}_i]$, participant i prefers to win the tournament because, as a winner, he/she gets a positive payoff for any $b_{i,e_j^{\text{com}}} \in (e_i^{\text{tangent}}, \hat{e}_i)$ and a zero payoff for $b_{i,e_j^{\text{com}}} = \hat{e}_i$ whereas participant i always get a zero payoff in case he/she rather decided to lose the tournament. QED

Claim. For any participant i 's belief such that $b_{i,e_j^{\text{com}}} > \hat{e}_i$, then participant i 's best response is to lose the tournament and exert $e_i^{\text{com}^L} = 0$.

Proof. With reference again to the Figure 3.A.5.1, we see that that in the range $(\hat{e}_i, +\infty)$ the total cost is always higher than the expected total revenue, which means that $e_i^{\text{com}^W} = b_{i,e_j^{\text{com}}} > \hat{e}_i$ always yields a negative payoff, whereas participant i always get a zero payoff in case he/she rather decided to lose the tournament. QED

Summing up, we can analytically represent participant i 's best response function as:

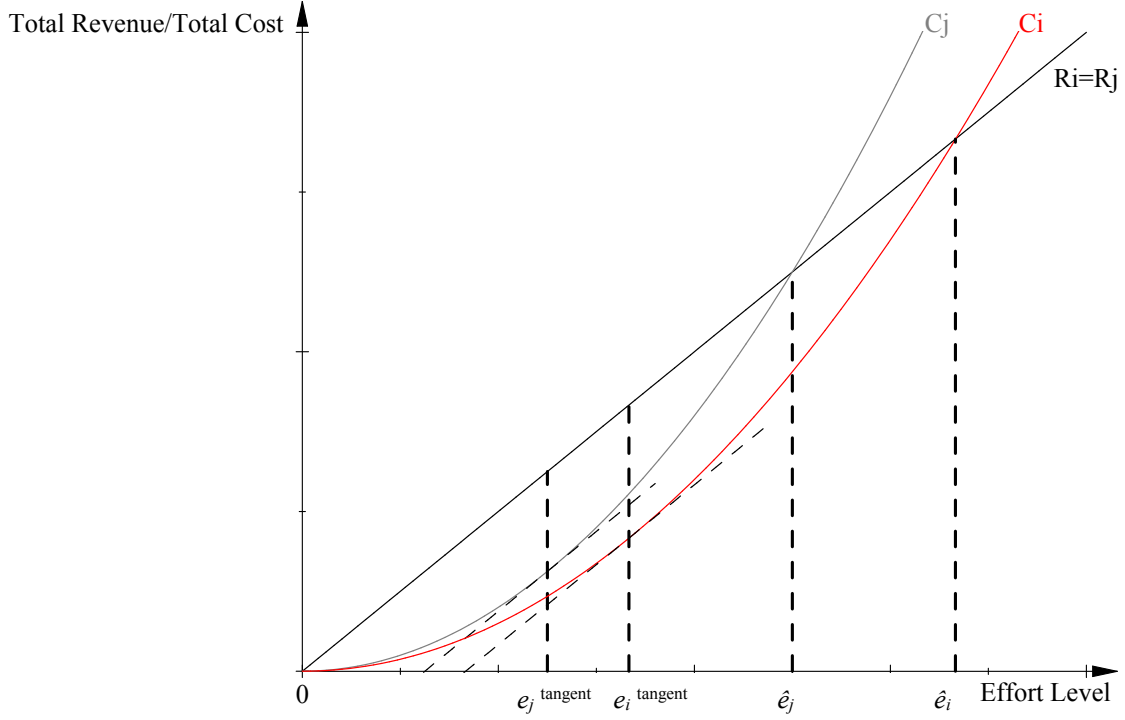
$$(2) \quad BR_i(b_{i,e_j}) = \begin{cases} e_i^{\text{com}^*} = e_i^{\text{tangent}} & \text{if } 0 \leq b_{i,e_j^{\text{com}}} \leq e_i^{\text{tangent}} \\ e_i^{\text{com}^*} = b_{i,e_j^{\text{com}}} & \text{if } e_i^{\text{tangent}} < b_{i,e_j^{\text{com}}} \leq \hat{e}_i \\ e_i^{\text{com}^*} = 0 & \text{if } b_{i,e_j^{\text{com}}} > \hat{e}_i \end{cases}$$

in which $e_i^{\text{com}^*}$ represents participant i 's optimal level of effort in the competitive second stage.

3.A.5.2. Asymmetric case I

In Figure 3.A.5.2, we depict qualitatively the Asymmetric case I:

Figure 3.A.5.2. Asymmetric case I



If we simultaneously observe Figure 3.A.5.2 and the best response function of participant i in the symmetric case, we conclude that we can analytically represent the best response function of participant i still with the analytical expression in (2).

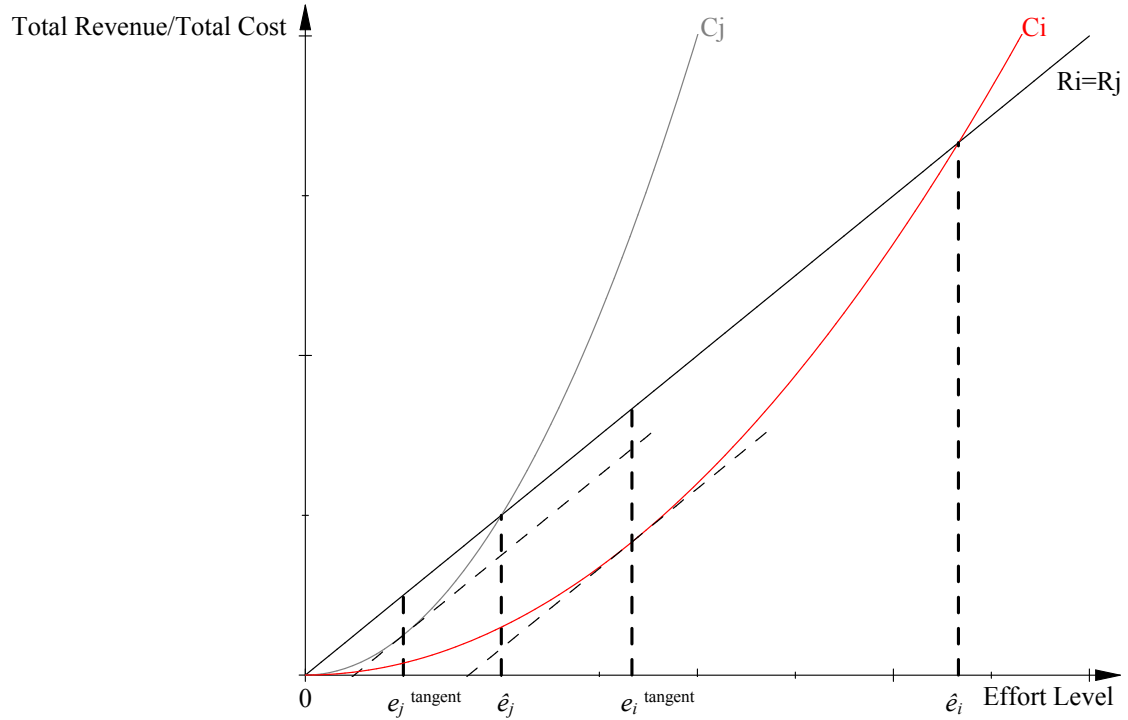
Differently to the Symmetric case, however, in the Asymmetric case I we have to take into consideration the existence of profitable deviations within $[e_j^{\text{tangent}}, e_i^{\text{tangent}})$ and within $(\hat{e}_j, \hat{e}_i]$. With reference to the Figure 3.A.5.2, we see that:

1. In the range $[e_j^{\text{tangent}}, e_i^{\text{tangent}})$, participant j does not deviate in any pair of strategies such that $e_i^{\text{com}} = e_j^{\text{com}}$, but participant i always deviates to e_i^{tangent} .
2. In the range $(\hat{e}_j, \hat{e}_i]$, participant i does not deviate in any pair of strategies such that $e_i^{\text{com}} = e_j^{\text{com}}$, but participant j always deviates to zero effort.
3. Regarding beliefs within the other ranges, the incentives to (or not to) deviate are analogous to the Symmetric case.

3.A.5.3. Asymmetric case II

In Figure 3.A.5.3, we depict qualitatively the Asymmetric case II:

Figure 3.A.5.3. Asymmetric case II



If we simultaneously observe Figure 3.A.5.3 and the best response function of participant i in the symmetric case, we again conclude that we can analytically represent the best response function of participant i still with the analytical expression in (2).

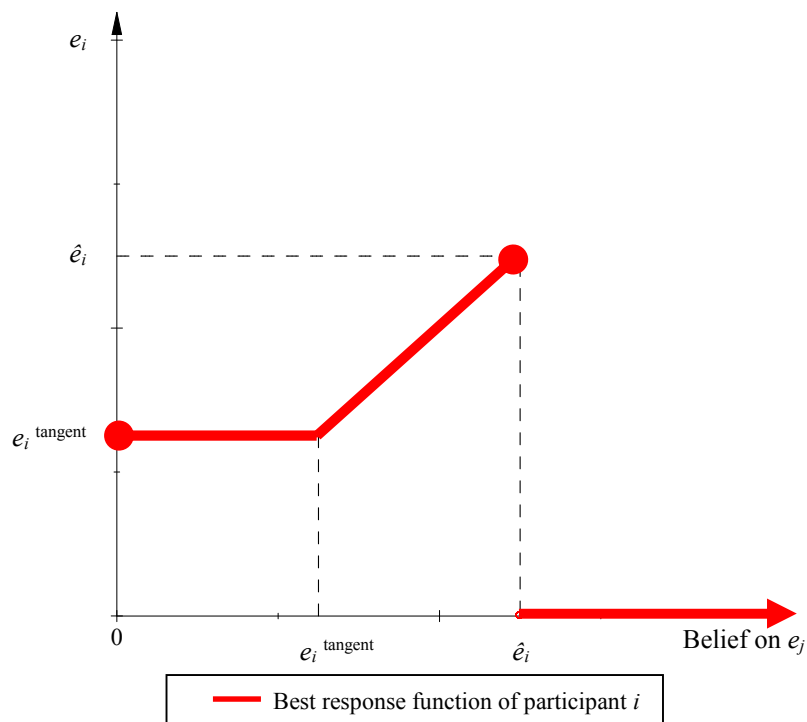
With reference to the Figure 3.A.5.3, we can trivially conclude that in this case participant i cannot rationally believe that $e_j^{\text{com}} > \hat{e}_j$ and participant j cannot rationally believe that $e_i^{\text{com}} \neq e_i^{\text{tangent}}$.

Appendix 3.A.6. Proofs of the propositions

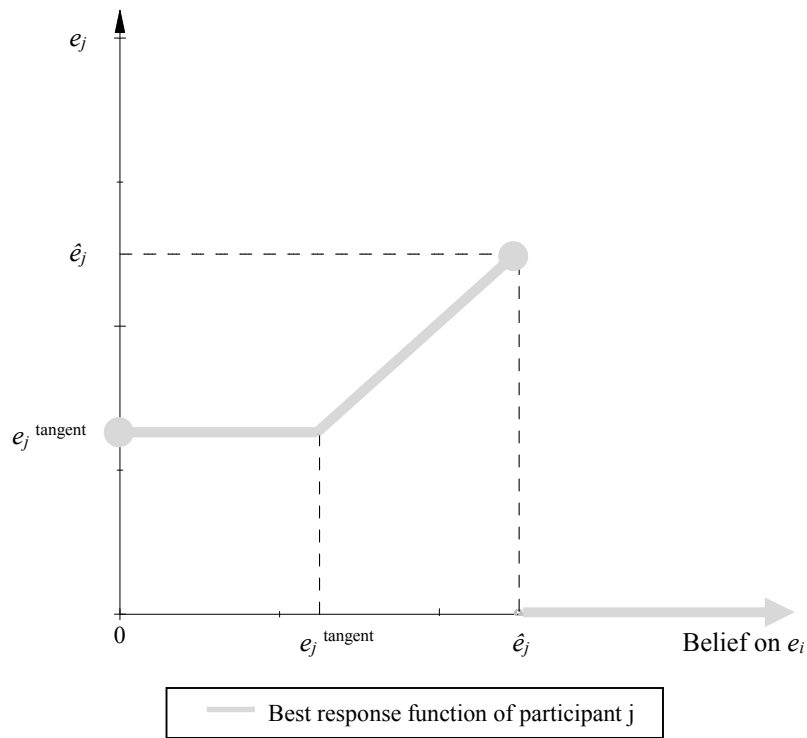
3.A.6.1. Proof of proposition 1

In any game, a Nash equilibrium corresponds to the intersection of the players' best-response functions.⁷ Thus, a simple way to demonstrate the set of Nash equilibria in the Symmetric case is to observe graphically the overlapping of the participants' best response functions that we derived analytically.

We first graph the best responses of participant i and participant j separately. Subsequently, we plot them together so that we observe where they overlap.



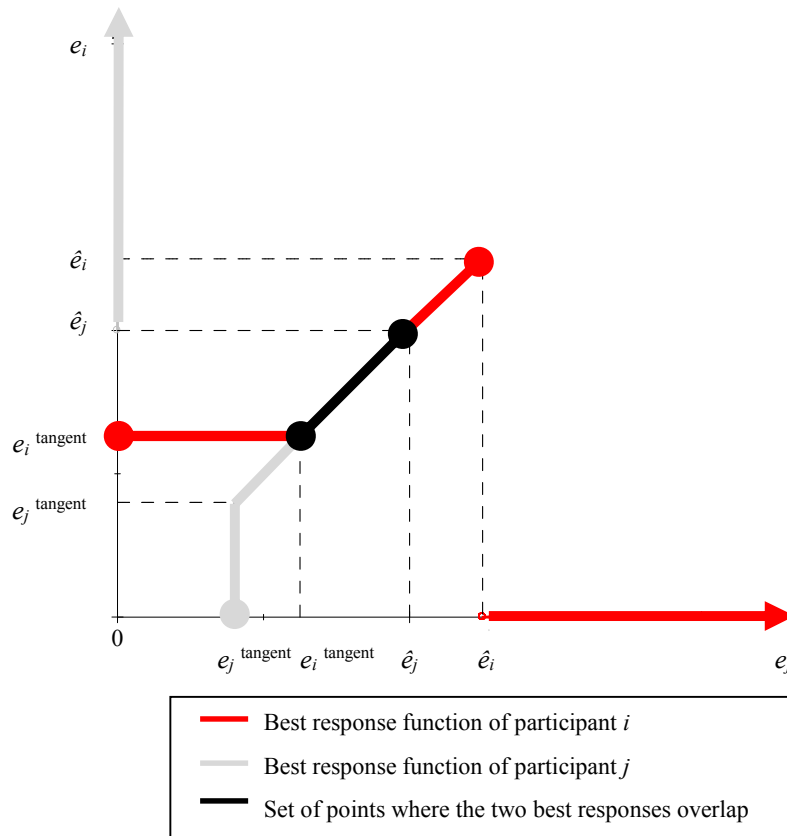
⁷ Or best response correspondences if the players' best response is a correspondence rather than a function.



With reference to the last graph, the points within the segment where the two best response functions overlap are, by definition of best response, the only points in the graph where both participants do not have an incentive to deviate. In other words, each point within the overlapping segment is a Nash equilibrium. We straightforwardly represent this set of Nash equilibria as *any pair* $(e_i^{com}, e_j^{com}) \in [e_i^{tangent}, \hat{e}_i]$ *such that* $e_i^{com} = e_j^{com}$. QED

3.A.6.2. Proof of proposition 2

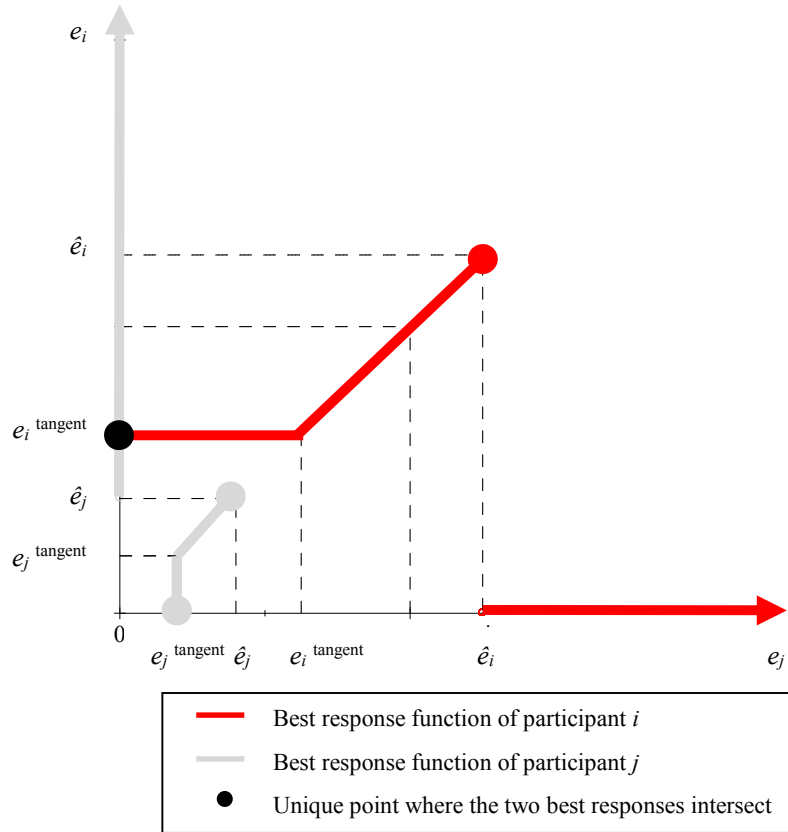
Following the same argument as in the proof of proposition 1, we plot together participant i 's and j 's best responses corresponding to the Asymmetric case I so that we observe where they overlap.



With reference to the graph above, the points within the segment where the two best response functions overlap are, by definition of best response, the only points in the graph where both participants do not have an incentive to deviate. In other words, each point within the overlapping segment is a Nash equilibrium. We straightforwardly represent this set of Nash equilibria as *any pair* $(e_i^{com}, e_j^{com}) \in [e_i^{tangent}, \hat{e}_j]$ *such that* $e_i^{com} = e_j^{com}$. QED

3.A.6.3. Proof of proposition 3

Again, following the same argument as in the proof of proposition 1, we plot together participant i 's and j 's best responses corresponding to the Asymmetric case II so that we observe where they overlap.



With reference to the graph above, we observe that there is a unique point where the two best responses intersect. Thus, by definition of best response, this is the unique point in the graph where both participants do not have an incentive to deviate, i.e., the unique point that is a Nash equilibrium. We straightforwardly represent this Nash equilibrium as $(e_i^{com}, e_j^{com}) = (e_i^{tangent}, 0)$.

QED

Appendix 3.A.7.

Economics literature on contests and tournaments⁸

3.A.7.1. A brief overview

To study the level of costly resources expended by participants that aim to win a contest or tournament, the economics literature has essentially been using three canonical models: i. The *Tullock contest* (e.g., Tullock 1980); ii. The *All-pay auction* (e.g., Hillman and Riley 1989); iii. The *Rank-order tournament* (e.g., Lazear and Rosen 1981). To distinguish the three approaches, we first introduce a general framework.

Consider a static game of complete information in which there are two risk-neutral participants that compete for a single prize of fixed value equal to v . Each participant, consider participant i for illustrative purpose, chooses e_i , i.e., how much effort to exert in the competition. Effort is costly and is represented by $c_i(e_i)$. The performance output of participant i , which is represented by y_i , depends on both participant i 's effort e_i and a random variable ε_i that is drawn from some common distribution with cumulative distribution F :

$$(1) \quad y_i = e_i + \varepsilon_i$$

in which the additive random noise ε_i can be thought as unobservable luck or performance error, for instance. In addition, participant i 's probability of winning the tournament depends on the observable output of both participants, $y = (y_i, y_j) \geq 0$. This relation is determined by the *tournament success function*:

$$(2) \quad p_i(y_i, y_j) = \begin{cases} \frac{y_i^r}{y_i^r + y_j^r} & \text{if } \max\{y_i, y_j\} > 0 \\ 1/2 & \text{otherwise} \end{cases}$$

in which $r \geq 0$ is a parameter that measures the sensitivity of the probability of winning with respect to a participant's output. The outcome-contingent payoff of participant i in the tournament is:

⁸ Two excellent surveys on this topic are provided in Konrad (2009) and in Dechenaux, Kovenock, and Sheremeta (2014). Our brief overview follows an exposition similar to Dechenaux et al. (2014).

$$(3) \quad \pi_i = \begin{cases} v - c_i(e_i) & \text{if } i \text{ wins} \\ -c_i(e_i) & \text{if } i \text{ loses} \end{cases}$$

Hence, participant i 's expected payoff is given by:

$$(4) \quad E(\pi_i) = p_i(y_i, y_j) \cdot v - c_i(e_i)$$

A simple version of a *Tullock contest* is obtained by setting $y_i = e_i$ in (1), $0 \leq r < \infty$ in (2), and $c_i(e_i) = e_i$ in (3). In this kind of tournament, the probability of winning the prize equals the ratio of individual effort to aggregate effort. For instance, in case $r = 1$, which has been called the *Lottery contest*, the general tournament success function in (2) becomes:

$$(5) \quad p_i(e_i, e_j) = \begin{cases} \frac{e_i}{e_i + e_j} & \text{if } \max\{e_i, e_j\} > 0 \\ 1/2 & \text{otherwise} \end{cases}$$

A simple version of the *All-pay auction* is obtained by setting $y_i = e_i$ in (1), $r = \infty$ in (2), and $c_i(e_i) = e_i$ in (3). In this case, the general tournament success function in (2) becomes:

$$(6) \quad p_i(e_i, e_j) = \begin{cases} 1 & \text{if } e_i > e_j \\ 0 & \text{if } e_i < e_j \end{cases}$$

The fundamental difference compared to the *Tullock contest* is that, ignoring the event of ties, the outcome is deterministic; the participant who exerts more effort wins the tournament with certainty.

Finally, to obtain a simple version of the *Rank-order tournament* we set $y_i = e_i + \varepsilon_i$ in (1), $r = \infty$ in (2), and $c_i(e_i) = c(e)$ in (3), in which $c'_i(e_i) > 0$, $c''_i(e_i) > 0$. The fundamental difference between the *Rank-order tournament* relatively to the *Tullock contest* and the *All-pay auction* is the noise component ε_i in the performance function (3). This implies that equilibria in this case depends on the distribution of the noise variable since the general tournament success function in (2) becomes:

$$(7) \quad p_i(y_i, y_j) = \begin{cases} 1 & \text{if } y_i > y_j \Leftrightarrow e_i - e_j > \varepsilon \\ 0 & \text{if } y_i < y_j \Leftrightarrow e_i - e_j < \varepsilon \end{cases} \quad \text{where } \varepsilon = \varepsilon_j - \varepsilon_i$$

3.A.7.2. Our framework and the literature

After describing the differences between the three canonical models, we would like to highlight a common feature to all of the approaches: the value of the prize is fixed ex-ante. In other words, the value of the prize v is exogenously determined.⁹

The competitive incentive scheme we use in the lab experiment consists of a winner-take-all tournament (w-t-a-t) with a specific feature that has not been theoretically studied hitherto. In contrast to a standard w-t-a-t, the value of the prize is not fixed ex-ante. It is rather endogenously determined by the winner's output.

Thus, assuming that the prize is fixed ex-ante is not appropriate to model behavior in the tournament of our lab experiment.¹⁰ The key distinction of our framework is therefore that we consider the value of the prize is dependent on a participant's output, i.e., $v(y_i)$. Moreover, we consider $r = \infty$, $y_i = e_i$ and $c'_i(e_i) > 0$; $c''_i(e_i) > 0$. This means that in our framework the tournament success function is given by (6) and participant i 's expected payoff in (4) becomes:

$$(8) \quad E(\pi_i) = p_i(e_i, e_j) \cdot v(e_i) - c_i(e_i)$$

In short, our framework assumes: i. The value of the prize is endogenously determined; ii. A tournament success function of an *All-pay auction* model; iii. Strictly convex costs as in a *Rank-order tournament* model.

⁹ The main motivation of the literature to conventionally consider the prize to be fixed ex-ante is the legal verifiability of a prize in court disputes.

¹⁰ Importantly, we would not be able to model the fact that a participant performing under the w-t-a-t that we use in the experiment could have an incentive to exert effort beyond the level at which a participant believes their effort is already sufficient to win the tournament.

3.A.7.3. Problems to incorporate incomplete information in our framework

To consider a static game of incomplete information is certainly more realistic to investigate the competitive stage of the lab experiment. A prominent theoretical analysis of an *All-pay auction* with convex costs and incomplete information is the model of Moldovanu and Sela (2001). As standard in the literature, their model assumes that the value of the prize is fixed ex-ante. If instead we assume that the value of the prize is endogenous, unfortunately severe analytical tractability problems arise.

We attempt two approaches to incorporate incomplete information in our framework. Assuming participants are privately informed about their own ability:

- i. In one approach, we assume the abilities are drawn independently from an interval according to a distribution function that is common knowledge. Accordingly, we attempt to maximize participant i 's expected payoff while guessing the functional form governing the relationship between e_j (the opponent's effort) and α_j (the opponent's ability) that is private information of the opponent. We tested one linear and two non-linear cases;
- ii. In the second approach, we apply a differential equation method to participant i 's payoff maximization problem f.o.c. in an attempt to retrieve a tractable functional form governing participant i 's probability of winning the tournament. In both approaches, we could not obtain a closed form solution. We are happy to share upon request the technical details of our attempts.

In brief, we are not successful to find a closed form solution when considering incomplete information. Although building a framework that considers a game of incomplete information to study a w-t-a-t in which the prize is endogenously determined by the winner's output is a promising avenue of research, further attempts to address this challenge go beyond the scope of this paper. Therefore, we use a framework that assumes a game of complete information to theoretically investigate the effort level participants might exert in the competitive stage of the lab experiment.

Part II

Group Membership

Group membership and individual competitive behavior¹

4.1. Introduction

The influence of group membership on individual decision-making is considered an undeniable fact in social science fields such as anthropology, political science, social psychology, and sociology. As elegantly discussed in Ellemers (2012), people who come together in groups change their sense of self. This phenomenon can lead to unwanted behavior. For instance, the aggression of a soccer hooligan might appear uncharacteristic when compared to same individual's temperament on his own. Conversely, the change of self in groups may also lead to desirable behavior such as the suppression of selfish actions, which is fundamental for inducing people to collaborate with each other.

Individual decision-making under group membership is manifestly present in numerous economically relevant settings including decisions within company boards, political races, sport competitions, and joint-work in a project. Thus, understanding how group membership alters individual behavior is necessary to prevent undesirable behaviors and/or promote better behavior in many economic contexts. Nevertheless, only recently economists have started challenging the standard assumption in economic theory that group membership should not alter individual behavior (for a seminal theoretical challenge, see Akerlof and Kranton 2000; for a seminal empirical challenge, see Charness, Rigotti, and Rustichini

¹ This chapter is written using the *pluralis majestatis* (also known as the royal “we”).

2007).² In this chapter, we investigate if (and how) group membership affects individual behavior in the particular context of a competitive setting.

We study two distinct aspects within a competitive setting. Firstly, we analyze individual choice of competition levels. Secondly, we assess individual performance under competitive incentives in front of an audience. The presence of an audience is common in economically relevant decision-making. For instance, at school or the workplace, one frequently performs competitive tasks under observation. Thus, we are also interested in understanding how being observed influences individual performance under competition, and if (and to what extent) group membership plays a role.

To test whether group membership affects individual competitive attitude, we conduct a controlled laboratory experiment. Our first aim for the experiment is to analyze participants' self-selection into a competition. Different from previous studies in experimental economics measuring competitiveness—which commonly use a binary indicator between a competitive and a non-competitive payment scheme to elicit willingness to compete (e.g., Niederle and Vesterlund 2007)—, we devise an indicator that allows measuring several different levels of competitiveness. Specifically, we ask each participant to choose among forty-nine different degrees of competition. The higher the competition level, the higher the winner's earnings and the lower the loser's earnings. In brief, our indicator ranges from a winner-take-all tournament payment scheme (maximum competition level) to nearly a piece-rate payment scheme (minimum competition level), in which the difference between the winning and losing payoffs is tiny. In other words, the higher the competition level, the steeper the incentives. Importantly, our measure of competitiveness rules out strategic considerations, which allows us to isolate the effect of competition better.

We elicit participants' choices of competition level in two different conditions. In the *Group Membership (GM) condition*, following classical social psychology experiments (see, e.g., Tajfel, Billig, Bundy, and Flament 1971; Tajfel and Turner 1979), we induce group membership using a Tajfel-style minimal group paradigm task. In it, we create two groups based on an assignment rule that is intended to be as meaningless as possible. To enhance group membership, we incentivize the feeling of being part of the group via payoff commonality, which means that individual behavior has a direct impact on the payoff of other members of the group. In the *Individual condition*, we do not induce group membership nor does individual behavior has an impact on the other participants' payoff. Moreover, in both conditions, we give feedback on the choices of competition level. Feedback is an element that—combined with the minimal group paradigm and payoff

² Other-regarding preferences, which were quite accepted in economics before 2000, already showed that individual behavior is not determined solely at the individual level. The novelty in the works of Akerlof and Kranton (2000) and Charness, Rigotti, and Rustichini (2007) is that they specifically study individual decision-making under group membership.

commonality—should amplify the saliency of the group in the GM condition (Charness et al. 2007).

The second aim of our experiment is to assess the performance of a real-effort task under competitive incentives in front of an audience. In the GM condition, the two groups are assigned to separate rooms so that each group member sequentially competes against a member of the other group via computer (one at a time). This means that while each participant is competing, an audience—composed of the other group members who are not competing—is silently observing the performance of the *active player*. The presence of a partisan audience should further amplify the salience of a group (Charness et al. 2007; Charness and Rustichini 2011). In the Individual condition, the setting is analogous to the GM condition. The important difference is that the audience is composed of individuals with no affinity to the active player.

Regarding the choices of competition level, our hypothesis is that salient group membership enhances individual competitive behavior. The basis for our hypothesis is three-fold. Firstly, experimental economics studies indicate that salient group membership increases aggressive behavior towards the outgroup. When inducing artificial group identities, studies show that salient group membership leads to targeting higher absolute material payoffs for the ingroup at the expense of the outgroup in different allocation tasks (Eckel and Grossman 2005; Charness et al. 2007; Chen and Li 2009; Chen and Chen 2011). Moreover, priming natural group identities, several studies show that group membership effects on behavior strengthen alongside the salience of group membership (Bernhard, Fehr, and Fischbacher 2006; Benjamin, Choi, and Strickland 2010; Goette, Huffman, Meier, and Sutter 2012; Cohn, Maréchal, and Noll 2015; Benjamin, Choi, and Fisher 2016). In our setting, aggressive behavior, framed as targeting higher absolute payoffs for the ingroup members, implies choosing a high competition level. Moreover, we purposely make group membership salient by incentivizing the feeling of GM via payoff commonality and by providing feedback on ingroup members choices of competition level.

Secondly, we consider that there is a noticeable risk attitude mechanism in our setting that might lead to higher choices of competition level in the GM condition. Specifically, the expected passive earnings in the GM condition—which do not exist in the Individual condition—might trigger risk-seeking behavior when a participant is choosing the competition level. In our setting, risk-seeking translates to choosing higher competition levels because that implies steeper incentives between the winning and the losing payoff. To test this mechanism, in the GM condition we elicit not only participants' (strict) individual risk attitude but also their individual risk attitude in a context where the individual risk-based decisions affect the payoff of the other group members.

Thirdly, assuming participants have other-regarding preferences, the participants' confidence in performing the real-effort task might differently affect

the willingness to compete in each condition because the choice of competition level has a positive externality on fellow group members in the GM condition. Following this line of reasoning, the precise connection between the choice of competition level and confidence depends on whether a participant is high- or low-confident. On the one hand, high-confident participants are more likely to believe that high competition levels lead to a stronger positive externality on other group members because their expected winning payoff is higher if they choose a high competition level. This means that high-confident participants might show bias for choosing a higher competition level compared to similarly high-confident participants in the Individual condition. On the other hand, for low-confident participants, the mechanism we suggest works in the opposite direction. We hypothesize that group membership turns participants more confident to compete because of *Teamgeist*, which means that the distribution of participants' confidence in the GM condition is "more to the right" relative to the Individual condition. Thus, the relative strength of the high-confident bias should be stronger in the GM condition. Accordingly, we expect higher choices of competition level in the GM condition through the confidence mechanism, *ceteris paribus*.

Concerning the individual performance of a competitive task while being observed, there seems not to exist an explanatory economic theory to predict the impact of group membership. Nevertheless, if we consider that participants in the GM condition feel support because of the partisan audience, we might expect group membership to enhance competitive performance.

Our results show that individuals who are members of a group choose lower levels of competition compared to individuals on their own. This difference is not explained by risk attitude nor by confidence. Hence, our hypotheses related to the choices of competition level are rejected. Interestingly, we find that second-order beliefs about the choices of competition level significantly predict participants' competitiveness in the GM condition, but not in the Individual condition. This finding accommodates a *guilt aversion* motive (see Charness and Dufwenberg 2006; Battigalli and Dufwenberg 2009), an explanation that we explore in the discussion of the results. In addition, we find that participants observed by their group mates while competing perform better than participants observed by "strangers". This result provides a thought-provoking insight into performance under competition before an audience, which we use to draw policy implications.

Our research relates to the growing literature in experimental economics that analyzes the impact of group membership on economic behavior. To illustrate the variety of applications, we briefly describe a few prominent studies. Charness et al. (2007) provide experimental evidence that salient group membership profoundly affects individual behavior in coordination and cooperation games. More specifically, they find that individuals who feel part of a group behave more aggressively compared to individuals who see themselves as isolated. Chen and Li (2009) show that the degree of social preferences in numerous allocation tasks is stronger towards members of an ingroup than towards members of an outgroup.

Goette, Huffman, Meier, and Sutter (2012) find that officers of the Swiss Army are more cooperative in a prisoner's dilemma game towards members of their own platoon compared to members of other platoons. Cohn, Maréchal, and Noll (2015) show that inmates from a maximum-security prison cheat more on a self-reporting task when their criminal identity is made salient. Benjamin, Choi, and Fisher (2016) find that making a person's religion salient affects contributions in a public goods game. Despite the diversity of contexts investigated in past research, hitherto the effect of group membership on individual competitive attitude remains unexplored.³ The central objective of this chapter is to unravel this connection.

Our study is also related to the economics literature on team decision-making (for an overview, see Cooper and Kagel 2005). The key difference between team decision-making and individual decision-making under group membership is that the former requires an agreement among team members before a decision is taken. Using an investment task, Sutter (2009) shows that individual decisions with salient group membership are, in the aggregate, identical to the decisions of unitary teams. Relevant to our study, Sutter's equivalence result is obtained in a non-strategic setting, which suggests that the results we report in this chapter are probably transferable to a team-decision making environment. Finally, our work can also be seen as a useful complement to social psychology research on social identity (Tajfel and Turner 1979) and self-categorization (Turner 1985).

In a nutshell, the central contribution of this study is two-fold. This is the first study in experimental economics to measure the impact of group membership on individual competitive attitude. Thus, we open a new avenue to advance our knowledge and policy advice regarding economically relevant competitive settings in which individual decisions are taken under group attachment. Second, on a methodological note, we create an indicator that measures several different degrees of competitiveness. Relatively to a standard binary indicator, our measure can capture more subtle differences in the sensitiveness to compete, which could be useful for researchers measuring competitiveness in a wide range of other contexts.

The remainder of the chapter is organized as follows. In section 4.2, we present the design of the experiment. In section 4.3, we present and discuss the choice of competition level results. In section 4.4, we present and discuss participants' performance of the real-effort task. In section 4.5, we conclude.

³ Perhaps one could consider close to our work the studies of Chowdhury, Jeon, and Ramalingam (2016) and Riedl and Zaunbrecher (2016) that investigate the connection between social identity and contests. However, their exclusive focus on Tullock contests means they analyze settings in which the use of resources/effort is wasteful from a social point of view, and there is no competitive performance of a real-effort task. Their research is therefore rather applicable to conflict and rent-seeking environments.

4.2. The experiment

To test if (and how) group membership influences individual competitive behavior, we conduct a laboratory experiment with two conditions: i. The *Group Membership (GM) condition* in which we induce participants to feel part of a group; ii. The *Individual condition*.

4.2.1. Experimental design

The experiment is divided into three parts. In the first and third parts, all participants are in experimental room 1. In the second part, participants are evenly split between experimental room 1 and experimental room 2. In the following, we describe the different steps of the experiment in the order they happen.

4.2.1.1. First part

In the first part of the experiment, all participants are seated in computer cubicles at experimental room 1.

- **General instructions.** We start by providing an overview of the experiment to the participants. We inform participants that the experiment consists of seven rounds. In two of these rounds, a participant has to perform a monetarily incentivized real-effort task. In five of these rounds, a participant only has to observe another participant performing the real-effort task. Following the order in which the rounds happen, we designate them as ROUND 0 to ROUND 6. Each round lasts 5 minutes. In addition, we ask questions before the start of ROUND 0 and ROUND 1, and after ROUND 6. Participants can earn additional money with their answers.

- **Risk preference start.** Right at the start of the experiment, we elicit each participant's risk preference using the measure of Dohmen, Falk, Huffman, and Sunde (2010). Specifically, participants make a decision in each of 15 binary choices between a fixed lottery and a sure amount. The fixed lottery yields € 4 with a 50% chance whereas the sure amount varies from € 0.25 (min) to € 3.75 (max) in increments of € 0.25 across the 15 decisions. By the end of the experiment, the computer randomly selects one decision situation for payment (see Appendix 4.A.3 for the full instructions).

- **Practice round.** Before ROUND 0, there is an unpaid practice round to familiarize the participants with the real-effort task. The task consists of adding up sets of five two-digit numbers. This addition task has been extensively used in experiments investigating competition (see, e.g., Niederle and Vesterlund 2007).⁴ Each participant practices the addition task for 2 minutes.

⁴ We do not allow participants to use a calculator, but they can use a pen and scratch paper. Participants face the same sequence of addition problems in all sessions. Participants read each problem on-screen and

▪ **ROUND 0: Addition task productivity.** Each participant solves as many addition problems as possible within 5 minutes under a piece-rate payment scheme that yields € 0.50 per correct answer.

▪ **Confidence.** Right after ROUND 0, we elicit each participant's confidence in performing the addition task (we describe this elicitation in detail in section 4.3.2).

▪ **Priming of group membership.** Evidently, only participants in the GM condition receive this priming. The method we use to induce group membership is based on Tajfel's minimal group-paradigm (Tajfel et al. 1971; Tajfel and Turner 1979). Specifically, we flash an image containing dots on the screen. Immediately after dots have been flashed, we ask each participant to guess how many dots were contained in the image (under no time limit). After participants provide their guesses, we form two groups—Group A and Group B—of six members, and inform each participant that they are grouped with five other participants “whose guess of dots is similar to yours”.⁵ Also, we inform participants that if they wish, they can inspect the allocation rule we use to form the two groups after the experiment. Finally, we inform participants that for ROUNDS 1–6, the six members of Group A will stay in the current experimental room, whereas the six members of Group B will move to another experimental room with one of the experimenters.

Regarding the Individual condition, at this moment we simply inform participants that for ROUNDS 1–6 we randomly assign six participants to stay in the current experimental room, whereas the other six participants will move to another experimental room with one of the experimenters.

4.2.1.2. Second part

For the sake of brevity, we take the GM condition to describe the remainder of the design. The key difference in the instructions of the Individual condition is that the words “member” and “group” in the instructions of the GM condition are replaced by the words “participant” and “room”, respectively (see Appendix 4.A.3 for the full instructions in both conditions).

The second part of the experiment consists of six competition rounds. In each round, one member of Group A competes against one member of Group B. In each group, we randomly assign each member to be the active player in one of the six rounds and to be a passive player in the five other rounds.

As the active player, a participant has two tasks: i. Choosing the competition level for the round in which a participant competes; ii. Performing the addition

must type the answer using the keyboard. Once an answer is submitted, the screen instantaneously displays a new problem.

⁵ The allocation rule we use to assign a participant to group A or B is whether a participant's guess is below or above the median guess (see Appendix 4.A.1 for the full details regarding the allocation rule).

task during 5 minutes under the chosen competition level against the active player of the other group.

As a passive player, a participant has two tasks: i. In each of the 5 rounds in which a participant is a passive player, a participant watches in silence the performance of the active player in the group; ii. After the end of the last competition round (i.e., after ROUND 6), pay attention to the feedback we provide on the choice of competition level made by each member of the group.

▪ **Choice of competition level.** In each experimental room, participants are first seated in separate computer cubicles to (privately and simultaneously) choose the competition level for the round in which they will be the active player. Specifically, we offer the following competitive payment scheme:

$$\begin{cases} \left(\frac{1}{2} + \frac{x}{100}\right) \times \text{€}1 & \text{per correct answer} & \text{if } \text{correct answers}_{\text{YOU}} > \text{correct answers}_{\text{OPPONENT}} \\ \left(\frac{1}{2} - \frac{x}{100}\right) \times \text{€}1 & \text{per correct answer} & \text{if } \text{correct answers}_{\text{YOU}} < \text{correct answers}_{\text{OPPONENT}} \end{cases}$$

for which each participant has to choose x belonging to $\{1, 2, 3, \dots, 49, 50\}$ to determine the exact competitive payment scheme. In case of a tie, a participant earns € 0.50 per correct answer regardless of his choice of x .⁶ We denominate this payment incentive scheme as the *Endogenous Tournament*. The x that we elicit is our measure of each participant's competitive attitude.

Importantly, we inform participants that: i. Each active player's earnings are calculated according to his chosen x . That is, the active player's competitive payment scheme does not depend on the x chosen by the opponent; ii. The active player does not learn the x chosen by his opponent neither before nor after the competition round.

Moreover, we add **payoff commonality** in the GM condition to enhance group membership. Specifically, just before participants choose the competition level, we add the following information to the instructions: “we would like to inform you that each of the other members of your group earns an amount equal to $1/3$ of your earnings in the round in which you compete. For example, if you earn € 12 in the competition, we not only pay you € 12 but also pay $1/3 \times \text{€ } 12 = \text{€ } 4$ to each of the other members of your group.”

⁶ E.g. 1, Choice of $x = 50$ (maximum competition level), which yields € 1 per correct answer if a winner, € 0 per correct answer if a loser, or € 0.50 per correct answer in case of a tie; E.g. 2, Choice of $x = 1$ (minimum competition level), which yields € 0.51 per correct answer if a winner, € 0.49 per correct answer if a loser, or € 0.50 per correct answer in case of a tie; E.g. 3, Choice of $x = 25$ (intermediate competition level), which yields € 0.75 per correct answer if a winner, € 0.25 per correct answer if a loser, or € 0.50 per correct answer in case of a tie. Each participant receives a payoff table that describes the winner/loser/tie payoff per correct answer corresponding to each $x \in \{1, 2, 3, \dots, 49, 50\}$ (see Appendix 4.A.1 for the full payoff table).

▪ ***Beliefs about the choice of competition level.*** Elicitation of each participant’s beliefs related to the choices of competition level (we describe this elicitation in detail in section 4.3.3). Participants are still seated in separate computer cubicles.

▪ ***ROUNDS 1–6: Competition.*** In each round, the active player sits in front of a computer to compete during 5 minutes against the active player of the other group. The five passive players merely observe in silence the active player during the competition. The passive players are seated behind the active player, but they cannot see the computer screen.

▪ ***Feedback on other’s choices of competition level.*** Right after ROUND 6 ends, we inform each participant about the competition level chosen by each of the other five members of his group.

4.2.1.3. Third part

In the third part of the experiment, all participants move back to their initial computer cubicle in experimental room 1.

▪ ***Risk preference end.*** We elicit each participant’s risk preference using again the measure of Dohmen et al. (2010). However, the elicitation in the GM condition has a twist so that we can measure possible differences in risk attitude when a participant’s risk-based decisions affect the payoffs of the other members of the group. Specifically, we elicit the same 15 binary choices of the start, but in the GM condition we add payoff commonality analogous to the one we use for the choice of competition level (we describe this elicitation in detail in section 4.3.2).

▪ ***Debriefing***

In Table 4.1, we highlight the key commonalities and differences between the GM and Individual conditions.

Table 4.1. Key features of the experimental conditions

	Guess of dots	Payoff commonality	Audience	Feedback
Group Membership	Yes	Yes	Yes	Yes
Individual	No	No	Yes	Yes

4.2.2. Experimental procedure

The experiment is computerized using z-tree software (Fischbacher 2007) and conducted in the SBE Computer Pool, part of the Behavioral and Experimental Economics Laboratory (BEElab) at Maastricht University. We recruit the participants via ORSEE software (Greiner 2015) from the subject pool of the BEElab. We conduct four sessions in each condition. Each session involves 12 participants. In total, 96 subjects participate in the experiment. To avoid possible confounds related to gender competition (see, e.g., Niederle and Vesterlund 2007), we exclusively invite male participants to participate in the experiment. Also, importantly for our purposes, participants predominantly do not know each other.⁷ The average age is 21. Each session lasts, on average, 120 minutes. Average earnings are € 22.95.

4.3. Competition level results

4.3.1. Choice of competition level

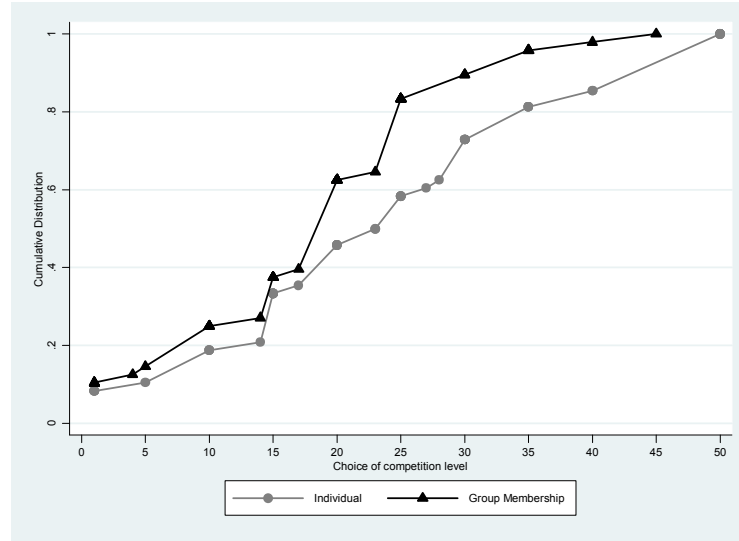
We report right away the choices of competition level. In the GM condition, participants choose, on average, a competition level of 19.13 whereas in the Individual condition participants choose, on average, a competition level of 24.81. These results reject our hypothesis that group membership enhances competitive behavior.

On the contrary, the results rather indicate that participants in the GM condition choose lower levels of competition [19.13 vs. 24.81, $n = 96$, $p = 0.069$, Mann-Whitney (MW) test]. We reach the same conclusion if we instead take each room or each session as a data point, respectively [room level cluster: $n = 16$, $p = 0.016$, MW test | session level cluster: $n = 8$, $p = 0.043$, MW test].

In Figure 4.1, we further show that the choice of competition level distribution in the Individual condition statistically dominates the choice of competition level distribution in the GM condition. Summing up, our first main result is:

Result 1: *When we induce group membership, participants choose lower competition levels. In other words, our measure of competitiveness indicates that group membership restrains individual willingness to compete.*

⁷ To the debriefing question: “How many of the participants in this experimental session do you consider as a person you know?”, participants report to know, on average, 0.15 and 0.33 persons in the GM and Individual conditions, respectively.

Figure 4.1. Cumulative distributions of the choice of competition level

4.3.2. Baseline performance, confidence, and risk attitude

Apart from the competitive attitude, participants' ability to perform the addition task, confidence in performing the addition task, and risk attitude are conspicuous factors that might influence their choice of the payment scheme in our setting.

To measure each participant's ability to perform the addition task, we use their performance in the piece-rate ROUND 0. To measure each participant's confidence in performing the addition task, we use their estimate of how many other participants (in percentage terms) correctly solved less problems than the participant himself. Specifically, right after ROUND 0, each participant answers the question: "All participants in this experiment have just performed ROUND 0 under the € 0.50 piece-rate payment scheme. How many of the other participants (in percentage) have correctly solved LESS addition problems than you did in ROUND 0?". The question is monetarily incentivized using a scoring rule that yields a higher reward for more accurate estimates (see Appendix A.4.3 for the full instructions, which provide the details of the monetary incentive scheme). To measure each participant's risk attitude (both at the start and at the end of the experiment), we consider their percentage of safe choices out of the 15 binary choices that we describe in section 4.2.1.1.

In Table 4.2, we report a summary of these measures. We observe that participants in the GM condition perform slightly worse and are slightly less confident, but the differences relatively to the Individual condition are not significant. With respect to the risk measures, we observe that participants in the GM are slightly less risk-averse at the start whereas they are slightly more risk-averse at the end of the experiment, but the differences relatively to the Individual condition are not significant either.

Table 4.2. Baseline performance, confidence, and risk attitude (in average)

Note: *Baseline performance* corresponds to the number of correct answers in ROUND 0. *Risk score* corresponds to the percentage of safe choices out of 15 binary choices between a sure amount and a lottery. *Confidence* and *Risk score* are variables that range from 0-100. p-values derived from Mann-Whitney tests of equality of distributions. Standard deviation in parentheses.

	Group Membership	Individual	Difference	p-value
Baseline performance	10.19 (4.7)	10.98 (5.0)	-0.79	0.423
Confidence	45.42 (25.5)	46.73 (30.0)	-1.31	0.823
Risk score at the start	50.55 (12.4)	51.25 (11.4)	-0.70	0.895
Risk score at the end	51.25 (11.6)	50.55 (11.0)	0.70	0.698

To measure possible differences in risk attitude when a participant takes into consideration his own group, recall that the risk measure that we elicit at the end of the experiment in the GM condition has a twist. Specifically, we elicit the same 15 binary risk-based decisions of the start, but each of the other members of the group earns an amount equal to $1/3$ of a participant's earnings. For example, if a participant earns € 3 in the randomly selected decision situation for payment, we not only pay € 3 to the participant but also $1/3 \times € 3 = € 1$ to each of the other participants in the group. The within-subject comparison between the two risk measures reveals no significant difference in the GM condition (50.55 vs. 51.25, $p = 0.326$, Wilcoxon signed-rank (WSR) test), which indicates that participants do not turn more or less risk averse when their choices affect the payoff of the other members of the group. In the Individual condition, as expected, the difference in the within-subject variation of risk attitude is not significant either (51.25 vs. 50.55, $p = 0.358$, WSR test).⁸

In Table 4.3, we present linear regression models of choice of competition level to assess the impact of these measures. The results in regressions (1) and (2) reveal that participants' ability to perform the addition task and confidence predict their choices of is, the higher is his choice of competition level. On the contrary, we observe that individual competition level in both the Individual and GM conditions. In particular, the more able and confident a participant risk preference does not significantly predict choices of competition level neither in the Individual condition nor the GM condition. Moreover, we observe in regression (2) that the change in the risk score in the GM condition does not significantly predict competitive choices either, which rejects *group risk attitude* as a mechanism through which GM could affect competitive behavior.

⁸ The results regarding risk attitude are qualitatively the same if we instead reported the shifting point from the lottery to the safe choice, in which case we would have to drop three participants from the analysis because of multiple shifting points.

Table 4.3. Linear regression models of choice of competition level I

Note: In regressions (3) and (4), participants in the Individual condition are the base group. *Confidence* and *Risk score* are variables that range from 0-100. *Risk score* refers to the risk score at the start of the experiment. *Change in risk score* equals (Risk score end - Risk score start). ***, **, and * significant at 1%, 5%, and 10%, respectively. Robust standard errors in brackets.

Dependent variable:	Choice of competition level			
	Individual (1)	Group membership (2)	Pooled (3)	Pooled (4)
<i>I if group membership</i>	n.a.	n.a.	-5.69** [2.582]	-4.78** [2.026]
<i>Baseline performance</i>	0.79* [0.427]	0.69* [0.365]		0.72*** [0.267]
<i>Confidence</i>	0.22*** [0.061]	0.15* [0.075]		0.19*** [0.047]
<i>Risk score</i>	0.003 [0.106]	-0.126 [0.080]		-0.069 [0.065]
<i>Change in risk score</i>	-0.033 [0.170]	-0.166 [0.152]		-0.101 [0.113]
Observations	48	48	96	96
R ²	0.393	0.406	0.049	0.417

The result in regression (3) of Table 4.3, in which we pool the two conditions, support our finding in section 4.3.1 that participants in the GM condition choose significantly lower levels of competition. Specifically, we see that participants induced with GM choose a competition level 5.69 lower compared to the participants in the Individual condition. Importantly, we show in regression (4) of Table 4.3 that the latter result is robust when controlling for individual baseline performance, confidence, risk score, and change in risk score. Therefore, our second main result is:

Result 2: *The treatment effect that we identify in section 4.3.1 is robust when controlling for individual baseline performance, confidence, and risk attitude. We, therefore, conclude that group membership indeed restrains individual willingness to compete in our setting.*

4.3.3. Beliefs about the choices of competition level

In this section, we further investigate why group membership restrains competition in our setting. To this end, we analyze the connection between participants' beliefs about choices of competition level and their actual choice of competition level.

We elicit two beliefs regarding the choices of competition level. Immediately after participants make their choice of competition level, we first ask participants

the following question in the GM (Individual) condition: “Consider the other members (participants) of your group (room). What is your estimate of their choice of competition level, on average?”. We denominate this belief as the *first-order belief*. Secondly, we ask participants the following question in the GM (Individual) condition: “Consider the other members (participants) of your group (room). What is your estimate of their average estimate of your choice of competition level?”. We denominate this belief as the *second-order belief*. The questions are monetarily incentivized using a scoring rule that yields a higher reward the more accurate the estimate is (see Appendix 4.A.3 for the full instructions, which provide the details of the monetary incentive scheme).

In Table 4.4, we observe that the first-order beliefs are not significantly different between the two conditions. We reach the same conclusion regarding the second-order beliefs. Moreover, the two beliefs have a significant Spearman rank correlation in each condition [GM condition: 25.63 (first-order beliefs) vs. 24.83 (second-order beliefs), $\rho = 0.754$, $p < 0.001$; Individual condition: 25.46 (first-order beliefs) vs. 24.06 (second-order beliefs), $\rho = 0.884$, $p < 0.001$]. Notably, however, the first- and second-order beliefs have a significant Spearman rank correlation with the choice of competition level only in the GM condition [GM condition: 25.63 (first-order beliefs) vs. 19.13 (competition level), $\rho = 0.333$, $p = 0.021$ | 24.83 (second-order beliefs) vs. 19.13 (competition level), $\rho = 0.441$, $p = 0.002$; Individual condition: 24.46 (first-order beliefs) vs. 24.81 (competition level), $\rho = -0.004$, $p = 0.977$ | 24.06 (second-order beliefs) vs. 24.81 (competition level), $\rho = -0.039$, $p = 0.792$].

Table 4.4. Beliefs about the choice of competition level (in average)

Note: *First-order beliefs* and *Second-order beliefs* are measures that can range from 1-50. p-values derived from Mann-Whitney tests of equality of distributions. Standard deviation in parentheses.

	Group Membership	Individual	Difference	p-value
First-order beliefs	25.63 (8.9)	25.46 (11.2)	0.17	0.871
Second-order beliefs	24.83 (7.9)	24.06 (11.2)	0.77	0.514

To further uncover this surprising connection between beliefs and choices of competition level in the GM conditions, we run a regression analysis. In Panel B of Table 4.5, we present augmented versions of the Table 4.3 linear regression models of choice of competition level to assess the impact of the second-order belief. In Panel A of Table 4.5, we again report the results of Table 4.3 for ease of comparison.

In regression (4), we see that the second-order belief about the choice of competition level does not significantly predict participants’ willingness to compete in the Individual condition. In stark contrast, we see in regression (5)

Table 4.5. Linear regression models of choice of competition level II

Note: In regressions (3) and (6), participants in the Individual condition are the base group. *Confidence* and *Risk score* are variables that range from 0-100. *Risk score* refers to the risk score at the start of the experiment. *Change in risk score* equals (Risk score end - Risk score start). In Panel A, we again report the results of Table 4.3 for ease of comparison. ***, **, and * significant at 1%, 5%, and 10%, respectively. Robust standard errors in brackets.

Dependent variable:	Choice of competition level					
	Panel A			Panel B		
	Individual (1)	Group membership (2)	Pooled (3)	Individual (4)	Group membership (5)	Pooled (6)
<i>I if group membership</i>	n.a.	n.a.	-4.78** [2.026]	n.a.	n.a.	-4.87** [2.035]
<i>Baseline performance</i>	0.79* [0.427]	0.69* [0.365]	0.72*** [0.267]	0.79* [0.427]	0.39 [0.362]	0.69** [0.271]
<i>Confidence</i>	0.22*** [0.061]	0.15** [0.075]	0.19*** [0.047]	0.22*** [0.062]	0.15** [0.062]	0.19*** [0.047]
<i>Risk score</i>	0.003 [0.106]	-0.126 [0.080]	-0.069 [0.065]	-0.002 [0.119]	-0.125 [0.076]	-0.057 [0.067]
<i>Change in risk score</i>	-0.033 [0.170]	-0.166 [0.152]	-0.101 [0.113]	-0.036 [0.171]	-0.164 [0.143]	-0.091 [0.114]
<i>Second-order belief</i>				-0.017 [0.181]	0.438*** [0.155]	0.090 [0.119]
Observations	48	48	96	48	48	96
R ²	0.393	0.406	0.417	0.394	0.499	0.421

that this regressor is highly significant in the GM condition, which means that participants' second-order belief about the choice of competition significantly influences their willingness to compete in the GM condition. Specifically, the higher (lower) is the competition level a participant believes the other members of his group expect him to choose, the higher (lower) is the competition level that a participant chooses, *ceteris paribus*.⁹ The latter result is robust when we instead cluster the standard errors at the room or session level.

In addition, comparing regressions (2) and (5), we observe that the impact of the second-order belief and the confidence in performing the addition task, respectively, on the choice of competition level are orthogonal. Finally, comparing regressions (3) and (6), in which we pool the two conditions, we see that when we

⁹ The first-order belief is significant when we regress regression (5) with this regressor instead of the second-order belief. But it becomes insignificant when we include both beliefs in regression (5). Thus, we drop the first-order belief since the results of regressions (4) – (6) are qualitatively the same if we also include the first-order belief as a regressor.

control for the second-order belief, the GM condition dummy keeps significant and its magnitude nearly the same. The latter result is not surprising since we already knew that the second-order beliefs are similar in both conditions.

In a nutshell, the second-order beliefs of participants significantly predict willingness to compete in the GM condition, but not in the Individual condition. These findings, combined with participant's lower willingness to compete in the GM condition, accommodate an explanation based on a guilt aversion motive (see, e.g., Charness and Dufwenberg 2006; Battigalli and Dufwenberg 2009). Guilt aversion means that an individual prefers to avoid taking a decision that the individual believes let others down (i.e., the individual would experience guilt if he takes the decision). This leads to a less standard concept of utility, in which an individual's preferences depend on his beliefs about the beliefs of others, even if there are no strategic considerations.

The key belief to test guilt aversion is an individual's second-order belief. As we report above, in the Individual condition—in which the “other” participants in the room are just “strangers”—we find no connection between the second-order beliefs and the choice of competition level. Contrastingly, in the GM condition—in which the “other” participants in the room are members of the same group—we find a significant and substantial positive connection between the second-order beliefs and the choices of competition level, which indicates that the participants in the GM condition care about the expectations of other in the group while choosing the competition level. Collectively, these results provide support for the interpretation that participants in the GM experience guilt aversion while they decide the degree of competitiveness.¹⁰ We summarize in our third main result:

Result 3: *In the GM condition, a participant's belief about his group members' expectations about his choice of competition level significantly predicts a participant's choice of competition level, whereas that is not the case in the Individual condition. These findings indicate that participants induced with GM experience guilt aversion, which could explain why group membership restrains individual willingness to compete in our setting.*

4.3.4. Self-reported motivation

In the debriefing, we ask participants the following question: “*What was (were) your reason(s) to choose the Competition Level x for the round in which you were an ACTIVE PLAYER?*”. Five researchers who are not aware of the purpose of the study independently classify the answers to this question in one (or more) of the following five types of motivation: i. Confidence; ii. Risk attitude; iii. Losing payoff matters. This means the participant chooses x so that the losing payoff is

¹⁰ A further argument to support this interpretation is that we provide feedback about the other group members choices of competition level at the end of the experiment (and participants are aware of this before they choose the competition level). As Dostoyevsky brilliantly illustrates through the fictional character of Raskolnikov, other's awareness of one's acts is not a necessary condition to experience guilt. Nevertheless, we consider that ex-ante awareness that other's will be aware of one's acts should amplify one's prospective experience of guilt.

high enough (in the participant's perspective); iv. Joy of competing and/or winning. This means the participant chooses x so that the participant feels more joy while competing and/or in case of winning the competition; v. Focus self-assurance. This means the participant chooses x as a commitment device to be more focused. We associate a motivation(s) to a given answer if at least three researchers associate that motivation(s) to the answer (see Appendix 4.A.2 for the detailed classifications of the five researchers).

In Table 4.6, we concisely describe participants' self-reported motivation (see Appendix 4.A.2 for the answer of each participant). The striking difference between the two conditions is the higher proportion of participants in the GM condition that reveal they care about the losing payoff. This difference in motivation in the GM condition is consistent with our interpretation of the results. In the GM condition, the second-order belief about the choice of competition level represents a participant's belief about what the other group members expect to be the minimum payoff rate that he should guarantee for them (i.e., the losing payoff for the chosen x). Accordingly, participants especially caring about the losing payoff in the GM condition is consistent with participants experiencing guilt aversion.

Table 4.6. Type of (self-reported) motivation for the choice of competition

	Confidence	Risk attitude	Losing payoff	Joy of competing	Focus self-assurance
Group membership	65%	19%	46%	0%	0%
Individual	71%	25%	17%	2%	4%

Under this motive, for a given second-order belief, a participant prefers not to deviate upward in his choice of the competition level because that implies a lower losing payoff. That is, the more a participant deviate upward, the higher is the negative impact on the payoff of the other members of his group in case a participant loses the competition. In short, for a guilt-averse participant, the other group members' expectations about his choice of competition level are an important basis for his actual choice of competition level.

Alternatively, we could argue that participants' particular inclination to care about the losing payoff in the GM condition is evidence of maximin preferences (see, e.g., Charness and Rabin 2002; Engelmann and Strobel 2004). Strictly speaking, the choice of competition level in our setting in line with maximin preferences is $x = 1$, which ensures the maximum losing payoff. Nevertheless, the significantly lower choice of x in the GM condition can be the result of a stronger individual maximin preference when participants' choices affect the payoff of the other members of the group. That is, holding constant other motives such as confidence, risk attitude, etc., the relatively stronger maximin preference of

participants in the GM condition bias downward the choice of x compared to the Individual condition.

4.3.5. Other-regarding preferences—discussion

Models of social preferences assume that an individual's utility function depends not only on his own material payoff but also on the allocation of resources within a reference group, i.e., an individual is concerned about the material resources other people receive.

Our experiment is not purposely designed to test a specific social preference. Nevertheless, since participants in the GM condition have a salient reference group and their choice of competition level affects the resources the other group members receive, we discuss the possible role of other-regarding preferences.

We find that the difference in participants' willingness to compete between the two conditions squares with both guilt aversion and maximin motives. On the one hand, we find a connection between the second-order beliefs about the choice of competition level and the actual choices of competition level, which supports a guilt aversion motive. On the other hand, participants' self-reported motivation for the choice of competition level accommodate a motive à la maximin. An important argument in favor of the latter motive is the absence of strategic considerations in our setting¹¹ since a maximin motive seems to lose relevance only in strategic settings (see, e.g., Güth and van Damme 1998; Okada and Riedl 2005). In short, both motives likely play a role in our setting, but our data do not allow us to disentangle them.

Without intending to be exhaustive, we further consider alternative other-regarding motives. Firstly, if we think of outcome-based models capturing inequality aversion as a motive, they cannot explain our data. When a participant is an active player in the GM condition, his choice of competition level (as well as his performance) influences the size of the “cake” to be distributed, but not the actual split of the “cake”. Secondly, the inherent uncertainty of our competitive setting implies that there is no choice of competition level that unambiguously generates the biggest possible “cake” among all possible choices. Nevertheless, efficiency concerns (in the sense of maximizing the sum of payoffs) can emerge if participants use their beliefs about their relative performance to calculate ex-ante efficiency. However, the connection between confidence and choice of competition level in our data does not support an efficiency motive. Finally, our design rules out, or at least patently casts doubt on the relevance of other-regarding motives such as altruism, fairness, envy, reciprocity, spitefulness, and trust.

¹¹ We purposely rule out strategic considerations in our setting by neither turning the choice of competition level in a coordination game nor informing participants about the opponent's choice of competition level before the competitive performance (which would allow the possibility of strategic signaling).

4.4. Real-effort task performance results

In this section, we analyze participants' performance of the addition task. As discussed in the introduction, we are interested in evaluating the impact of being observed on individual performance of a competitive task, and whether group membership plays a role in this connection.

Recall that in both conditions each participant performs the addition task firstly under a non-competitive payment scheme and secondly under a competitive payment scheme. Thus, if we consider the (generally accepted) assumption that competitive incentives lead to a boost in performance, it is reasonable to hypothesize an increase in the number of correct answers in the competition round relatively to ROUND 0 in both conditions. Moreover, belonging to a group might amplify the latter effect because of the presence of an audience during the competition, which in the GM condition is composed of the other members of the group. That is, if participants feel more support due to the partisan audience, this could lead to an extra boost in their performance.

4.4.1. Accuracy in the addition task

In the GM condition, the average number of correct answers is 10.19 in the non-competitive ROUND 0 and 11 in the competitive round. In Table 4.7, we see that this difference is significantly different, which means participants perform better in the competitive round. In addition, comparing the cumulative distributions of performance of both rounds corroborates that participant's performance increases under competition (see Appendix 4.A.2). In the Individual condition, the picture is different. The average number of correct answers is 10.98 in the non-competitive ROUND 0 and 11.02 in the competitive round. In Table 4.7, we see that this difference is not significant, which means that participants do not significantly improve their performance in the competitive round. Regression analysis corroborates these two results (see Appendix 4.A.2). We summarize in our fourth main result:

Result 4: *Participants significantly increase the number of correct answers under competition only when induced with group membership.*

Table 4.7. Correct answers (in average)

Note: p-values derived from Mann-Whitney tests of equality of distributions. Standard deviation in parentheses.

	ROUND 0	Competition	Change in correct answers	p-value
Group Membership	10.19 (4.7)	11 (4.8)	0.81	0.019
Individual	10.98 (5.0)	11.02 (4.9)	0.04	0.529

4.4.2. Effort provision

Regression analysis also shows that in each condition the change in correct answers from ROUND 0 to the competition is not significantly affected by choice of competition level (see Appendix 4.A.2). This latter result should not be a surprise given the competitive incentive scheme that we offer. Recall that participants perform the competition round under the following *Endogenous Tournament* incentive scheme:

$$\left\{ \begin{array}{ll} \left(\frac{1}{2} + \frac{x}{100} \right) \times \text{€}1 & \text{per correct answer} \quad \text{if} \quad \boxed{\text{correct answers}_{YOU} > \text{correct answers}_{OPPONENT}} \\ \left(\frac{1}{2} - \frac{x}{100} \right) \times \text{€}1 & \text{per correct answer} \quad \text{if} \quad \boxed{\text{correct answers}_{YOU} < \text{correct answers}_{OPPONENT}} \end{array} \right.$$

That is, independently of a participant's choice of x , for a participant to become the winner a participant must correctly solve more addition problems than his opponent (note the part that we highlight with dashed line).

Hence, we not only expect participants to exert higher effort in the competition compared to the non-competitive ROUND 0 because of the competitive incentives, but we also expect participants' effort provision in the competition not to depend on their choice of competition level.¹² In short, we consider that it is rational for participants to provide their best effort regardless of their choice of x .

Table 4.8. Attempts (in average)

Note: p-values derived from Mann-Whitney tests of equality of distributions. Standard deviation in parentheses.

	ROUND 0	Competition	Change in attempts	p-value
Group Membership	12.58 (4.5)	13.92 (4.7)	1.34	< 0.001
Individual	13.29 (4.8)	14.67 (4.6)	1.38	< 0.001

Following Gerales, Riedl, and Strobel (2017), we use participants' number of attempts to tackle the addition problems as a proxy measure of effort provision. In Table 4.8, we see that participants in the GM condition significantly increase their number of attempts to solve the addition problems in the competitive round. We reach the same conclusion when considering the participants in the Individual condition. In brief, participants try harder in the competition relatively to ROUND 0 in both conditions. Regression analysis supports the latter two results and also shows that in each condition the change in attempts from ROUND 0 to

¹² One could argue that we could expect more effort with higher x because incentives become steeper. We consider this is not the case because: i. The steepness of the incentives is not exogenously set; ii. Steeper incentives do not necessarily induce higher effort provision (Huck, Kübler, and Weibull 2012)

the competition is not significantly driven by the choice of competition level (see Appendix 4.A.2).

To gain more insight, we further analyze participants' error rate, which equals—for each participant—the number of incorrect answers divided by the number of attempts to solve addition problems. In Table 4.9, we observe that the error rate of participants in the GM condition is virtually the same across the two rounds, whereas the participants' error rate in the Individual condition significantly increases in the competition round. Regression analysis corroborates these two results. And it also shows that the choice of competition level does not predict the error rate in neither condition (see Appendix 4.A.2).

Table 4.9. Error rate (in average)

Note: p-values derived from Mann-Whitney tests of equality of distributions. Standard deviation in parentheses.

	ROUND 0	Competition	Change in error rate	p-value
Group Membership	0.22 (0.16)	0.23 (0.15)	0.01	0.505
Individual	0.2 (0.15)	0.27 (0.15)	0.07	0.005

Thus, our final main result is:

Result 5: *Participants induced with group membership significantly improve performance during the competition because they keep the same accuracy rate while attempting more problems. In contrast, participants on their own are not able to improve their performance under competition. Even though they also attempt more problems, there is a significant decrease in their accuracy rate.*

4.4.3. Presence of an audience

The reason for the improvement in performance under competition only in the GM condition is open for speculation. All in all, we interpret the results in sections 4.4.1 and 4.4.2 in light of the presence of an audience during the competition rounds. In both conditions, when a participant is competing there is an audience observing the participant's performance. However, there is a crucial difference. In the GM condition, the audience is composed of the other members of the group, whereas in the Individual condition the audience is composed of "strangers". Therefore, we conjecture that the participants in the GM condition improve performance under competition because they feel supported while being observed by other group members. On the contrary, participants in the Individual condition do not improve performance under competition because they feel tense while being observed by "strangers". That is, the presence of "strangers" in the audience causes a participant to feel anxious during the competitive performance,

which leads him to make more mistakes while trying to solve more addition problems.

Evidently, it would be pertinent to have asked participants debriefing questions about their competitive performance. Unfortunately, this has not been the case, which means that our data do not allow us to test our conjecture regarding participants' perception of the audience during the competitive performance. Thus, future research to investigate in more detail the role of an audience in competitive settings is important since the presence of an audience is common in many economically relevant contexts.

4.5. Concluding remarks

This study contributes to a growing body of literature that integrates social identity with economic analysis. Specifically, the present research is the first attempt in experimental economics to assess the impact of group membership on individual competitive behavior.

The main goal of the current study was to experimentally investigate the impact of group membership on individual degree of competitiveness. We find that participants induced with group membership choose significantly lower levels of competition compared to participants on their own. Also, we notably find that, for the participants induced with group membership, second-order beliefs about competitiveness support a guilt aversion motive for the participants' actual choice of competition level. The choice of competition level in our setting affects the payoff of others only when a participant is a member of a group. Thus, our results indicate that participants induced with group membership make more cautious choices of competition level because they believe the other members of the group expect them to ensure "a high enough" payoff.

Naturally, more research is needed to evaluate the robustness of guilt aversion as a mediator motive linking group membership to competitiveness. Nevertheless, if this motive proves strong in other settings, our findings suggest an interesting policy implication to deter reckless decisions in competitive labor market: Index the bonus of workers to the performance of their peers. This incentive scheme could be applied, for instance, to tackle the infamous problem of irresponsible investment in the hedge fund world.

The second aim of this study was to analyze whether the presence of an audience affects individual performance of a real-effort task under competitive incentives. We find that being observed by members of one's group leads to better performance compared to the case when the audience is composed of individuals without a specific affinity with the performer. This finding suggests that in competitive environments in which individuals often work in open office space (e.g., in investment banking) strengthening the group spirit to compete

against outside firms rather than promoting internal competition better motivates workers.

One possible alternative explanation for our results is loss aversion (Kahneman and Tversky 1984) in combination with other-regarding concerns. Assuming individuals are loss-averse, if while choosing the competition level an individual considers the potential losses versus potential benefits that his choice has on the payoff of the other members of the group, then group membership should amplify loss aversion. Thus, this mechanism also implies lower choices of competition level for individuals induced with group membership relative to individuals on their own.

This research yields several questions in need of further investigation. Firstly, the reader should be informed that this study is part of an incomplete research program. As a follow-up step, we will conduct two additional conditions in which we vary the degree of salience of group membership. In one condition, we aim to test whether the mere fact of being part of a group is sufficient to lower the choice of competition level, or (as our present results suggest) whether payoff commonality is fundamental to restrain individual competitive behavior. In a second condition, we aim to remove feedback on the choices of competition level. This will allow us testing whether ex-ante awareness that other's will be aware of one's choice is a necessary condition to observe individual weaker inclination to compete when being part of a group. If removing feedback from our setting would wipe out the restraining effect of group membership on the choices of competition level, this would mean that loss aversion cannot explain the main result of the present study.

In closing, this study offers an experimental setting for future studies investigating the connection between group membership and competitiveness. More broadly speaking, a number of extensions deserve to be considered. Firstly, it would be worthwhile to add a strategic dimension to our setting. To this end, a natural extension to our setting would be to turn the choice of competition level into a coordination game. A second interesting extension to our setting would be to elicit the choice of competition sequentially. That is, an individual would observe the choices of preceding participants before taking his turn to choose the competition level. This would allow for an analysis of the effect of feedback on choices dynamically. Finally, our findings also indicate a significant interaction between the presence of an audience, group membership, and competitive performance. The thought-provoking behavioral implications of this result to competitive labor markets discussed above demand further research to examine more closely the links between these three elements.

REFERENCES

- Akerlof, G., & Kranton, R. (2000). Economics and Identity. *Quarterly Journal of Economics*, 115(3), 715-53.
- Battigalli, P., & Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144(1), 1-35.
- Benjamin, D., Choi, J., & Fisher, G. (2016). Religious identity and economic behavior. *Review of Economics and Statistics*, 98(4), 617-37.
- Benjamin, D., Choi, J., & Strickland, J. (2010). Social Identity and Preferences. *American Economic Review*, 100(4), 1913-28.
- Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group Affiliation and Altruistic Norm Enforcement. *American Economic Review*, 96(2), 217-21.
- Charness, G., & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, 74(6), 1579-1601.
- Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, 117(3), 817-69.
- Charness, G., & Rustichini, A. (2011). Gender differences in cooperation with group membership. *Games and Economic Behavior*, 72(1), 77-85.
- Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual Behavior and Group Membership . *American Economic Review*, 97(4), 1340-52.
- Chen, R., & Chen, Y. (2011). The Potential of Social Identity for Equilibrium Selection. *American Economic Review*, 101(6), 2562-89.
- Chen, Y., & Li, S. (2009). Group Identity and Social Preferences. *American Economic Review*, 99(1), 431-57.
- Chowdhury, S., Jeon, J., & Ramalingam, A. (2016). Identity and group conflict. *European Economic Review*, 90, 107-21.
- Cohn, A., Maréchal, M., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies*, 82(4), 1289-1308.
- Cooper, D., & Kagel, J. (2005). Are two heads better than one? Team versus individual play in signaling games. *The American economic review*, 95(3), 477-509.
- Dohmen, T., Falk, A., Huffman, H., & Sunde, U. (2010). Are Risk Aversion and Impatience Related to Cognitive Ability. *American Economic Review*, 100(3), 1238-60.
- Eckel, C., & Grossman, P. (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization*, 58(3), 371-92.
- Ellemers, N. (2012). The Group Self. *Science*, 336(6083), 848-52.
- Engelmann, D., & Strobel, M. (2004). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *American Economic Review*, 94(4), 857-69.

- Fischbacher, U. (2007). z-Tree Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10(2), 171–78.
- Geraldes, D., Riedl, A., & Strobel, M. (2017). Sex and performance under competition: Is there a stereotype threat shadow? An intriguing behavior of men. *Mimeo*.
- Goette, L., Huffman, D., Meier, S., & Sutter, M. (2012). Competition between organizational groups: Its impact on altruistic and anti-social motivations. *Management Science*, 58(5), 948-60.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-25.
- Güth, W., & van Damme, E. (1998). Information, strategic behavior and fairness in ultimatum bargaining: an experimental study. *Journal of Mathematical Psychology*, 42(2-3), 227–47.
- Huck, S., Kübler, D., & Wiebull, J. (2012). Social norms and economic incentives in firms. *Journal of Economic Behavior & Organization*, 83(2), 173-85.
- Kahneman, D., & Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*, 39(4), 341-50.
- Leibbrandt, A., Gneezy, U., List, & J. (2013). Rise and fall of competitiveness in individualistic and collectivistic societies. *Proceedings of the National Academy of Sciences*, 110(23), 9305-08.
- Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics*, 122(3), 1067-1101.
- Okada, A., & Riedl, A. (2005). Inefficiency and social exclusion in a coalition formation game: Experimental evidence. *Games and Economic Behavior*, 50(2), 278–311.
- Riedl, A., & Zaunbrecher, H. (2016). Social identity and group contests. *mimeo*.
- Sutter, M. (2009). Individual behavior and group membership: Comment. *The American Economic Review*, 99(5), 2247-57.
- Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In *The social psychology of intergroup relations* (pp. 33-47). ed. S. Worchel; Austin, W.. Monterrey CA: Brooks/Cole.
- Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*, 1(2), 149-78.

Appendix 4.A.1.

Additional notes on the experimental procedure

4.A.1.1. Allocation rule to evenly assign participants between two groups

The allocation rule we use to assign a participant to group A or B is whether his guess of the number of dots is below or above the median guess.

Formally, we use the following rule:

$$\begin{cases} \text{If } Estimate_i < \text{median estimate, then participant } i \text{ is assigned to Group A} \\ \text{If } Estimate_i > \text{median estimate, then participant } i \text{ is assigned to Group B} \end{cases}$$

In which $Estimate_i = \lceil Actual\ Estimate_i - 0.01 + (random_i * 0.02) \rceil$, $Actual\ Estimate_i$ refers to participant i 's estimate of dots, $random_i$ returns a uniformly distributed number between 0 and 1 for participant i , and $i = 1, 2, \dots, 12$.

We consider $Estimate_i = \lceil Actual\ Estimate_i - 0.01 + (random_i * 0.02) \rceil$ rather than $Actual\ Estimate_i$ to prevent the formation of uneven groups. That is, with this procedure we ensure that in case two (or more) of the participants have an estimate exactly equal to the median (something that did not happen), the groups we form still consist of six members each.

4.A.1.2. Payoff table for the choice of competition level

Competition Level	Payoff per correct answer		
	WINNER	LOSER	In case of a TIE
$x = 1$	€ 0,51	€ 0,49	€ 0,50
$x = 2$	€ 0,52	€ 0,48	€ 0,50
$x = 3$	€ 0,53	€ 0,47	€ 0,50
$x = 4$	€ 0,54	€ 0,46	€ 0,50
$x = 5$	€ 0,55	€ 0,45	€ 0,50
$x = 6$	€ 0,56	€ 0,44	€ 0,50
$x = 7$	€ 0,57	€ 0,43	€ 0,50
$x = 8$	€ 0,58	€ 0,42	€ 0,50
$x = 9$	€ 0,59	€ 0,41	€ 0,50
$x = 10$	€ 0,60	€ 0,40	€ 0,50
$x = 11$	€ 0,61	€ 0,39	€ 0,50
$x = 12$	€ 0,62	€ 0,38	€ 0,50
$x = 13$	€ 0,63	€ 0,37	€ 0,50
$x = 14$	€ 0,64	€ 0,36	€ 0,50
$x = 15$	€ 0,65	€ 0,35	€ 0,50
$x = 16$	€ 0,66	€ 0,34	€ 0,50
$x = 17$	€ 0,67	€ 0,33	€ 0,50
$x = 18$	€ 0,68	€ 0,32	€ 0,50
$x = 19$	€ 0,69	€ 0,31	€ 0,50
$x = 20$	€ 0,70	€ 0,30	€ 0,50
$x = 21$	€ 0,71	€ 0,29	€ 0,50
$x = 22$	€ 0,72	€ 0,28	€ 0,50
$x = 23$	€ 0,73	€ 0,27	€ 0,50
$x = 24$	€ 0,74	€ 0,26	€ 0,50
$x = 25$	€ 0,75	€ 0,25	€ 0,50
$x = 26$	€ 0,76	€ 0,24	€ 0,50
$x = 27$	€ 0,77	€ 0,23	€ 0,50
$x = 28$	€ 0,78	€ 0,22	€ 0,50
$x = 29$	€ 0,79	€ 0,21	€ 0,50
$x = 30$	€ 0,80	€ 0,20	€ 0,50
$x = 31$	€ 0,81	€ 0,19	€ 0,50
$x = 32$	€ 0,82	€ 0,18	€ 0,50
$x = 33$	€ 0,83	€ 0,17	€ 0,50
$x = 34$	€ 0,84	€ 0,16	€ 0,50
$x = 35$	€ 0,85	€ 0,15	€ 0,50
$x = 36$	€ 0,86	€ 0,14	€ 0,50
$x = 37$	€ 0,87	€ 0,13	€ 0,50

$x = 38$	€ 0,88	€ 0,12	€ 0,50
$x = 39$	€ 0,89	€ 0,11	€ 0,50
$x = 40$	€ 0,90	€ 0,10	€ 0,50
$x = 41$	€ 0,91	€ 0,09	€ 0,50
$x = 42$	€ 0,92	€ 0,08	€ 0,50
$x = 43$	€ 0,93	€ 0,07	€ 0,50
$x = 44$	€ 0,94	€ 0,06	€ 0,50
$x = 45$	€ 0,95	€ 0,05	€ 0,50
$x = 46$	€ 0,96	€ 0,04	€ 0,50
$x = 47$	€ 0,97	€ 0,03	€ 0,50
$x = 48$	€ 0,98	€ 0,02	€ 0,50
$x = 49$	€ 0,99	€ 0,01	€ 0,50
$x = 50$	€ 1	€ 0	€ 0,50

Appendix 4.A.2. Additional data

4.A.2.1. Self-reported motivations—GM condition

Note: Five researchers who are not aware of the purpose of the study independently classify the answers in one (or more) of the following five types of motivation: i. C = Confidence; ii. R = Risk attitude; iii. L = Losing payoff matters; iv. J = Joy of competing and/or winning; v. F = Focus self-assurance. We associate a motivation(s) to a given answer if at least three researchers associate that motivation(s) to the answer. According to this criterion, we could not associate into any motivation one answer. We report the answers *Sic Erat Scriptum*, i.e., they include typos, grammatical errors, etc.

ID	Answer	C	R	L	J	F
1	Didn't want to choose as high as 50 but confident enough to go half way.	<u>5</u>	1			
2	I worked on the basis of 0.50c per correct question. The added value of beating the other player I gave 0.20c and therefore the decreased value fo losing to the other player was also 0.20c.			<u>3</u>	1	
3	According to my own competetiveness I chose 20 due to my lacking math skills. However, there was probably another person in the other group that was also lacking skills in calculating such tasks. Therefore, I chose the level 20.	<u>5</u>	1		1	
4	I felt very confident in my abilities, however, the danger to lose and letting down my group mates made me chose an average competition level.	<u>5</u>	<u>3</u>	<u>4</u>		
5	i knew I was not too bad in additionning, so i was confident i had a chance to win	<u>5</u>				
6	After I thought I did not perform too well in the test round, I fine-tuned my system of calculating additions with two digit numbers that enhanced my performance. i now used the pen and paper provided. that gave me the confidence to be able to beat my oponent and win more money. hence performance level 30.	<u>5</u>				
7	I was confident that I could beat my opponent, but still wanted to ensure a payoff for my group in case I lose.	<u>5</u>		<u>5</u>		
8	I choose the competition level based on my qualitties of math and the earnings. I quit good in math, so I choose to be at a higher level than 25 but then I realized the earnings and I also thought the other competitor could be really good so I chose 20 to ensure a higher minimal payoff for my group.	<u>5</u>	2	<u>5</u>		
9	It was determined by level of low risk		<u>5</u>			
10	I thought that there is a good chance that my opponent is better than me and I wanted to get paid for my answers at least a minimum of 40.	<u>4</u>		<u>5</u>		
11	I felt I would be below average, so I minimized risk to a certain extent.	<u>4</u>	<u>5</u>			

ID	Answer	C	R	L	J	F
12	i expected myself to be better then 3/5 of the group (according to my previous estimation of who was better/worse than I in the testing phase). so i set my competition level appropriatate = 3/5 of 50x = 30x	<u>5</u>				
13	because i thought i was a good compromise in case i lose or in case i win, i ll still received sufficient money, in the first case		1	<u>5</u>	1	
14	i was not sure enough about winning then to choose 50. However, I was sure about beeing not so bad that I would definately loose, so i did not choose 1. In the end i thought 23 is a good thing in between.	<u>5</u>	1	1		
15	Make sure to receive a decent amount even if I would be the losing competitor		1	<u>5</u>		
16	I thought I was fairly good at mental arithmetic, so had a reasonable chance of earning a high amount.	<u>5</u>				
17	Due to my own experiences with my analytical and academic level, also compared to others, I assumed that my performance will be above the average. This led to the assumption that the chance of answering more questions right than the average opposite player are quite high. Therefore, I choosed a competition level above 75% (equal to the percentage I have assumed in the question before the start of the \"joined\" round, after round 0.	<u>5</u>				
18	I wanted a small compensation in case I am the better player, i.e. the winner, therefore I chose it not be only 1. But I was afraid loosing all my gains, I decided to stick with a relatively small amount to reduce the risk		<u>5</u>	<u>5</u>	1	
19	I chose the competitiopn level because I wanted to take a risk to maximize my profits, however I also chose the level bearing in mind that my failure shuoldn't have too much of a negative activity on my fellow teammates.	1	<u>5</u>	<u>5</u>		
20	For me the possible payoff in case of winning was worth the risk of almost getting nothing if I loose and that is why I chose a high competitive level		<u>5</u>		1	
21	try to find a balance between winning a high amount and receiving at least some money when loosing		2	<u>5</u>	1	
22	I think I'm not very good in doing such calculations. So I expected to loose.	<u>5</u>				
23	I chose x=1, because i was quite certain that I was going to lose the competition.	<u>5</u>				
24	enough money in case i win, not all lost in case i loose		2	<u>5</u>	1	
25	because I was quite confident to win (which I dont know yet) but I also wanted to have some payoff when I lose. and I knew that I will answer at least 10 correct ones	<u>5</u>		<u>5</u>		
26	Knowledge of my low ability compared to the likely ability of other students.	<u>5</u>				
27	I could get a high level of correct answers. But stay safe in case I lose	<u>3</u>	<u>3</u>	<u>4</u>		
28	Maximum payoff in case of winning, while still receiving some payoff in case of losing. A balance between the two.	1	2	<u>5</u>	1	

ID	Answer	C	R	L	J	F
29	Tradeoff between win and loss.	1	2	<u>3</u>	1	
30	I chose the Competition Level X = 10 in the round in which I was an active player because I'm not really good at computing under a time pressure, it's why i tried to earn anyways something	<u>5</u>		1		
31	I think many players were better than me, but I was stil somehow confidet to win so I chose 10.	<u>5</u>				
32	Cause i am not that good in calculating in my head. So I thought that the propability to win will be very small!	<u>5</u>				
33	As in my opinion the # of my correct answers in round 0 was that low, i did not wanted a low payoff due to my loss.	2		<u>5</u>		
34	The main reason consisted of my physical wellbeing. Competition level 1 gave the best of the worst choices.			<u>4</u>		
35	i felt not as i could keep up with a high level because of my low calculation skills	<u>5</u>				
36	based on the result of my first round, where I just had 3 answers in total correct.	<u>5</u>				
37	my previus knowledge suggest my that my math level was not sufficient to try the test on a really high level of uncertaty	<u>5</u>	<u>3</u>			
38	I thought I was above average. Howver, I specified a number that would ensure a reasonabe payoff to my group in case i lose	<u>5</u>	1	<u>5</u>		
39	I think I am below the average.	<u>5</u>				
40	I wanted to have a better payment for a good answer if I win than if I lose, so I choose a higher x - but i didnt want it to be 0 in case I lose...	1		<u>4</u>	2	
41	Confidence in winning.	<u>5</u>				
42	Mainly the payoff division of 0,65 and 0,35, as it seemed as a good one to the group			2		
43	I choose 25 because I had the feeling the job I was doing in Round 0 was on an intermediate level and I could hold up that level during the competition. With 25 you win more if you win and don't loose everything if you lose		<u>5</u>			<u>5</u>
44	I think Im one of the \"better' players, therefore i chose 30. But that was actually stupid. should have taken 50 instead.		<u>5</u>			
45	i just assumed my chances to loose would be high.		<u>5</u>	1		
46	I wanted some security in case I am competing against a real fast thinker.			<u>3</u>	<u>4</u>	
47	I chose level 20 because whilst winning could mean a decent profit at this level, losing won't be too bad.			1	<u>5</u>	1
48	I've always been told that I'm good at mental calculation. So it was very unlikely that I get an opponent that is better than me.		<u>5</u>			

4.A.2.2. Self-reported motivations—Individual condition

Note: Five researchers who are not aware of the purpose of the study independently classify the answers in one (or more) of the following five types of motivation: i. C = Confidence; ii. R = Risk attitude; iii. L = Losing payoff matters; iv. J = Joy of competing and/or winning; v. F = Focus self-assurance. We associate a motivation(s) to a given answer if at least three researchers associate that motivation(s) to the answer. According to this criterion, we could not associate into any motivation one answer. We report the answers *Sic Erat Scriptum*, i.e., they include typos, grammatical errors, etc.

ID	Answer	C	R	L	J	F
1	I was quite sure that my chances to win the round would not be too high, therefore I chose the safest option for the case that I lose. However, if I won, the amount of money is still enough in my opinion.	<u>5</u>	1	<u>4</u>	1	
2	I trusted my mathematical skills but i didn't want to go all the way to 50	<u>5</u>				
3	There was chance that someone could beat me, so payment 2:1 (win:lose) was naturally acceptable.	<u>4</u>	2	1		
4	I knew I am not a math crack and would be likely to lose so I went for least risk and highest outcome.	<u>4</u>	<u>5</u>	1		
5	My reason to choose the Competition level of 23 was that one the one hand I am confident that there are players performing better and worse than I do, so I picked something in the middle. On the other hand I went another two level down to 23, because I thought that the probability that I am performing worse than my opponent player is higher than 50 %. So I choosed a little bit less than the middle.	<u>5</u>	1	1		
6	I was confident that I beat my opponent	<u>5</u>				
7	At first I wanted to play safe and go only with x=10. In hindsight I regret choosing 20.	1	2	2		
8	I thought that I was better in calculating the given problems than 75 - 80 % of the people. I therefore chose the given payout of 80 : 20.	<u>5</u>				
9	confidence let me choose a level of 30. Let me see if it paid off :)	<u>5</u>				
10	I felt somewhat confident that I would win from the other active player, but not confident enough to go all the way to 50. Additionally the payments below 25 were too low in my opinion. 30 was the best trade off in my case.	<u>5</u>		1	1	
11	just gut feeling nothing specific	1				
12	I was ready to take the risk since I wa confident in having many correct answers	<u>5</u>	<u>4</u>			
13	An option in the middle, but a bit higher than the middle.	2		1		
14	How good I was in comparison to other and not wanting to risk too much.	<u>5</u>	<u>5</u>			
15	Giving myself a little extra motivation, but not throwing away all the money when I would not win,		1	2	1	<u>5</u>
16	I thought that by doing so I could have earned more money.	1			1	1
17	Because I think the chance was not too high that my opponent is better	<u>5</u>				

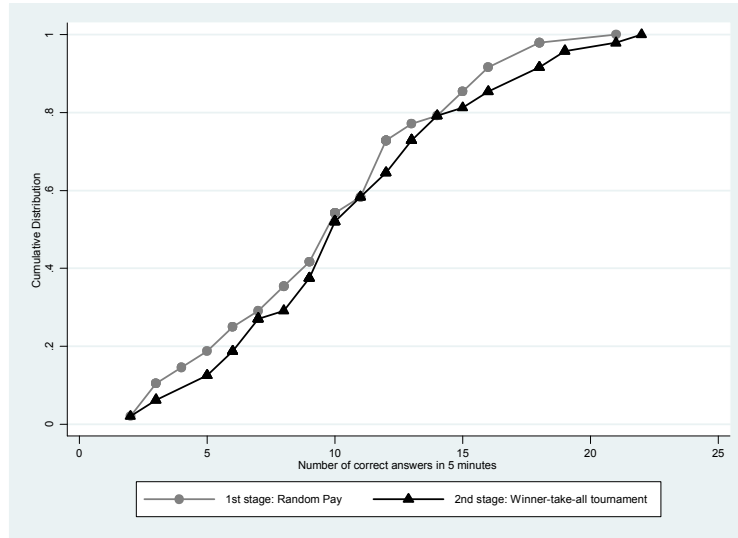
ID	Answer	C	R	L	J	F
18	Based on how well I would do relative to other players and concluded that i would probably lose so chose a low value for x.	<u>5</u>		1		
19	Well, I know that Im not good at maths. So my chance of winning is very slim.	<u>5</u>		1		
20	The probability of going up against a player stronger than me. I suspected i have a rather good chance of being better, therefore chose a rather high level.	<u>5</u>				
21	I chose the maximum because I thought it would be more fun, and possibly sharpen my concentration.				<u>5</u>	<u>5</u>
22	I had confidence in my skills to solve the equations	<u>5</u>				
23	I did not choose too high, because i want a certain kind of certainty. I chose not too low, because Round 0 went well and I thought the chance was there I would be smarter than my competitor	<u>4</u>	<u>3</u>	2		
24	I chose the competition level which would provide me with the minimum earnings that I was happy with in the event that I lost		1	<u>5</u>		
25	i thought i could win so i choose a number in which my earnings would be more then 51 however i was not sure to win so i also wanted to have money in case i lost.	<u>4</u>	1	<u>5</u>		
26	good split		2	1		
27	I performed good in the round before. I thought let's just take the risk.	<u>4</u>	<u>4</u>			
28	bad at math. others are certainly better than me. therefore low level.	<u>5</u>				
29	I picked level 15 because I wanted to secure a high enough payment based on my performance. Since I answered 9 questions correctly in round 0, I assumed that I would be able to answer around 10 questions correctly in the competition round as well. Therefore, if I won, I would earn 6.50 euros; however, if I lost, I would earn 3.50 euros, both of which are \"high enough\" payments for me.	2	1	<u>5</u>	1	
30	Taking into account both risk and award, I went for a rather \"save\" strategy. Also, you can never know against who you play and it was difficult to guess the other players' ability to calculate in a fast manner. Therefore, it is wise to go for a low X-value if you want to be save.		<u>5</u>	2		
31	I estimated that 25% answered less questions than I did. Hence, I tried to maximize my outcome by taking approximately the highest expected value X.	<u>4</u>	2	1		
32	My calculation skills are usually quite good, so it was no problem for me to take a risk.	<u>5</u>	<u>4</u>			
33	The x chosen depends on my own perception of my mathematical abilities and the people surrounding me in the experiment.	<u>5</u>				
34	general estimate of my own abilities > estimated likelihood that I would win, however with a decent margin that I still get some profit even if I play against a strong player and lose	<u>5</u>		<u>4</u>		

ID	Answer	C	R	L	J	F
35	I had the feeling that my performance in round 0 was above average, so there was a certain probability I would win the competitive round. Still, I wanted to have a sure payment of an adequate level.	<u>5</u>		<u>4</u>		
36	I chose on the basis of the estimation of my skills compared to the other players.	<u>5</u>				
37	My main reason for choosing competition level 50 was my confidence my abilities to calculate quickly and work well under pressure	<u>5</u>				
38	didn't want to take too much or too less risks so picked in between 1 and 50		<u>5</u>			
39	Because I am satisfied with both payoffs. The win would be deserved and earnings for loss are good enough			<u>5</u>	2	
40	Comparison of my expected performance against the expected performance of the other. I perceived myself a little bit above average, and I am not too risk loving in general so I picked 10	<u>5</u>	<u>5</u>			
41	1. I thought I did relatively well in the first round. 2. That was the only way to make real cash. Exposing myself to the risk was that not that big of a deal because the amount forgone by losing was rather low.	<u>4</u>	<u>5</u>	1	1	
42	I wanted to spread the risk of not earning any money hence I chose X=15 Especially since the practice round and the piece round did not go well for me I wanted to make sure that I play risk averse	<u>4</u>	<u>5</u>			
43	I was very confident that I could "beat" most other players, so I choose 50.	<u>5</u>				
44	To some extent balanced outcome as well as an acceptable level of risk.		<u>5</u>	1		
45	Well, there is always someone smarter than you. You never know who you are going against, it could be someone smarter or dumber, but I think I am above average. So I chose Competition level 30.	<u>5</u>	1			
46	I had the feeling my performance was good enough to match those of the other players. Hence, the X of 25.	<u>5</u>				
47	I wanted to ensure at least some income, partly independent of my performance.		1	<u>4</u>	1	
48	The way I think about myself and my knowledge	<u>5</u>				

4.A.2.3. Cumulative distributions of performance in the GM condition

Figure 4.A.2.3. Cumulative distributions (CDs) of performance in the GM condition

Note: The performance distribution in the competitive second round statistically dominates the performance distribution in the non-competitive first round. Additionally, we also run the Kolmogorov-Smirnov test, which yield $p < 0.001$.



4.A.2.4. Linear regression models of change in performance from the non-competitive to the competitive round

Table 4.A.2.4. Linear regression models of change in performance from the non-competitive to the competitive round

Note: In regressions (1) and (2), the regressor *Baseline performance* corresponds to the number of correct answers in ROUND 0. In regressions (3) and (4), the regressor *Baseline performance* corresponds to the number of attempts in ROUND 0. In regressions (5) and (6), the regressor *Baseline performance* corresponds to the error rate in ROUND 0. In all regressions, we demean each regressor so that the interpretation of the intercept is meaningful (e.g., $Confidence_i = \text{actual } Confidence_i - \text{sample mean of confidence}$). Specifically, in regressions (1) and (2), the intercept measures the change in the number of correct answers, for a participant with a baseline performance, confidence, risk score, change in risk score, and choice of competition level equal to the sample mean. The interpretation of the intercept in the other regressions is analogous. ***, **, and * significant at 1%, 5%, and 10%, respectively. Robust standard errors in brackets.

Dependent variable:	<i>Change in correct answers</i>		<i>Change in attempts</i>		<i>Change in error rate</i>	
	GM (1)	Individual (2)	GM (3)	Individual (4)	GM (5)	Individual (6)
<i>Baseline performance</i>	-0.08 [0.089]	-0.21** [0.097]	0.010 [0.076]	-0.019** [0.087]	-0.53*** [0.150]	-0.58*** [0.124]
<i>Confidence</i>	0.006 [0.019]	-0.016 [0.020]	0.001 [0.015]	-0.011 [0.013]	-0.001 [0.001]	0.0001 [0.001]
<i>Risk score</i>	0.02 [0.038]	0.02 [0.031]	0.02 [0.020]	0.04* [0.022]	-0.0003 [0.002]	0.0002 [0.002]
<i>Change in risk score</i>	0.03 [0.046]	-0.03 [0.051]	0.07** [0.027]	-0.01 [0.043]	0.003 [0.003]	0.002 [0.003]
<i>Choice of competition level</i>	0.001 [0.035]	0.025 [0.050]	0.005 [0.029]	0.040 [0.032]	0.001 [0.002]	-0.0002 [0.002]
<i>Intercept</i>	0.81** [0.320]	0.4 [0.421]	1.34*** [0.202]	1.36*** [0.296]	0.013 [0.018]	0.075*** [0.020]
Observations	48	48	48	48	48	48
R ²	0.04	0.148	0.154	0.193	0.324	0.304

Appendix 4.A.3.

Experimental instructions

4.A.3.1 Experimental instructions—First part

Note: After the welcome speech, participants sit in private computer cubicles at experimental room 1.

Welcome speech [*Given by one experimenter in the welcome room before participants move into computers cubicles*]

Good morning/afternoon! I welcome you to this experiment. My name is Diogo Geraldes, and I am the organizer of today's experiment. This experiment is part of a research project on decision-making financed by a scientific institution. The approximate duration of the experiment is 2 hours. All the information provided in this experiment is truthful. If you wish, you can inspect this at the end of the experiment.

In the lab, you can earn money with the decisions you make. The amount you earn will be paid out in cash, individually and confidentially, at the end of the experiment.

On the card you are holding, you see a number. In the lab, the computer cubicles are also numbered. You will take a seat in the computer cubicle with the same number as your card.

I ask you now a moment before you enter the computer room (quick check of z-leafs and number of participants)

You are about to enter the computer room and take a seat in the computer cubicle that corresponds to your number. We use these numbers for the data analysis. The number will not be linked to your name. Your anonymity is therefore secured.

In the lab, you will receive detailed instructions about the experiment. It is for us important that you understand and follow the instructions carefully. In particular, I ask you not to communicate in any other way than specified in the instructions.

Before you enter the lab, please switch off your mobile phone and store it inside your bag/pocket.

Are there any questions at this moment? Please, you can enter now into the computer room.

Welcome screen [*same version in the two conditions*]

Welcome to the SBE COMPUTER POOL!

You will now take part in a decision-making experiment.

Please click on <continue>.

<Continue>

General instructions [*same version in the two conditions*]

You are now taking part in an experiment in decision making. For showing up on time today you will be paid € 3. In addition, you can earn money with the decisions you make. Hence, it is important that you fully understand the instructions that follow. Please read them carefully.

In this experiment, you will experience seven rounds. In two of these rounds, you have to perform a monetarily incentivized work task (we describe how your earnings are determined before the start of the rounds). In five of these rounds, you only have to observe in silence another participant performing the work task. Following the order in which the rounds will happen, we designate them as: ROUND 0, ROUND 1, ROUND 2, ROUND 3, ROUND 4, ROUND 5, and ROUND 6. Each round will take 5 minutes.

Moreover, we will ask you some questions during the experiment. You can earn additional money with your answers.

Before you experience ROUND 0, there is an unpaid PRACTICE ROUND to make you familiar with the work task.

If you understand these instructions, please click on the OK-button. Otherwise, please raise your hand. An experimenter will then come to you and answer your questions.

<OK>

Instructions for the elicitation of risk preference—start [*same version in the two conditions*]

Before reading the PRACTICE ROUND instructions, please read the following.

In the following screen, you will have to make a choice between a lottery and a sure payment in 15 decision situations. The lottery is the same across the 15 decision situations, but the sure payment is always different.

Once you have made a choice for each one of the 15 decision situations, the computer randomly selects one decision situation for payment. Your earnings will depend on your choice for that decision situation. If you have chosen the lottery for that decision situation, you will earn with a 50 percent chance 400 points and with a

50 percent chance 0 points. If you have chosen the sure payment for that decision situation, you will earn the sure payment corresponding to that decision situation.

The points you earn in this part of the experiment will be exchanged into Euros at the end of the experiment with an exchange rate of:

100 points = € 0.50

If you have a question, please raise your hand.

Please click CONTINUE to make your choices.

<CONTINUE>

Elicitation of risk preference—start [*same version in the two conditions*]

Please decide for each decision situation whether you want to choose the lottery or the sure payment!			
Decision Situation	Lottery	Sure payment	Your choice
1	50% chance of 400 points and 50% chance of 0 points	25 points	lottery <input type="radio"/> sure payment <input type="radio"/>
2	50% chance of 400 points and 50% chance of 0 points	50 points	lottery <input type="radio"/> sure payment <input type="radio"/>
3	50% chance of 400 points and 50% chance of 0 points	75 points	lottery <input type="radio"/> sure payment <input type="radio"/>
4	50% chance of 400 points and 50% chance of 0 points	100 points	lottery <input type="radio"/> sure payment <input type="radio"/>
5	50% chance of 400 points and 50% chance of 0 points	125 points	lottery <input type="radio"/> sure payment <input type="radio"/>
6	50% chance of 400 points and 50% chance of 0 points	150 points	lottery <input type="radio"/> sure payment <input type="radio"/>
7	50% chance of 400 points and 50% chance of 0 points	175 points	lottery <input type="radio"/> sure payment <input type="radio"/>
8	50% chance of 400 points and 50% chance of 0 points	200 points	lottery <input type="radio"/> sure payment <input type="radio"/>
9	50% chance of 400 points and 50% chance of 0 points	225 points	lottery <input type="radio"/> sure payment <input type="radio"/>
10	50% chance of 400 points and 50% chance of 0 points	250 points	lottery <input type="radio"/> sure payment <input type="radio"/>
11	50% chance of 400 points and 50% chance of 0 points	275 points	lottery <input type="radio"/> sure payment <input type="radio"/>
12	50% chance of 400 points and 50% chance of 0 points	300 points	lottery <input type="radio"/> sure payment <input type="radio"/>
13	50% chance of 400 points and 50% chance of 0 points	325 points	lottery <input type="radio"/> sure payment <input type="radio"/>
14	50% chance of 400 points and 50% chance of 0 points	350 points	lottery <input type="radio"/> sure payment <input type="radio"/>
15	50% chance of 400 points and 50% chance of 0 points	375 points	lottery <input type="radio"/> sure payment <input type="radio"/>

Please click OK if you have made your choice for each decision situation!

OK

Practice round [*same version in the two conditions*]

This PRACTICE ROUND serves to familiarize you with the work task with which you can earn money in subsequent ROUNDS. In this PRACTICE ROUND, you are not paid. However, it is in your best interest to take this round seriously because the more practice you gain now, the more money you are likely to earn in the two rounds in which you have to perform the work task.

In this round, you will be asked to calculate the sum of 5 two-digit numbers (e.g.: $84 + 94 + 17 + 19 + 39 = ?$). You will be given 2 minutes to calculate as many of these addition problems as possible.

Please use the keyboard to type in your answer into the box and click the "OK"-button to confirm it. The computer will immediately tell you whether your answer

was correct or wrong and a new addition problem will instantaneously be displayed on your screen. Your answers to the problems are anonymous.

You must answer the addition problems without any helping device such as a pocket calculator or a mobile phone. You are welcome, though, to make use of the provided scratch paper in case you wish to write the numbers down.

The remaining time in seconds will be shown on the top right of the screen. If you have a question, please raise your hand.

Please click on the OK-button to start the PRACTICE ROUND.

<OK>

Round 0—Piece-rate [same version in the two conditions]

In this round, you will be given 5 minutes to calculate as many addition problems as possible. The procedure to submit your answers is the same as before.

In this round, you get € 0.50 per problem that you correctly solve. Incorrectly solved problems do not count (neither positively nor negatively). We refer to this payment as the *piece-rate* payment.

Example 1: If you correctly solve 35 addition problems in the 5 minutes, then you will earn: € 17.5 = 35 x € 0.50.

Example 2: If you correctly solve 4 addition problems in the 5 minutes, then you will earn: € 2 = 4 x € 0.50.

The numbers in the examples are just for illustrative purposes. They DO NOT intend to suggest how anyone should perform in this experiment.

The remaining time in seconds will be shown on the top right of the screen. If you have a question, please raise your hand.

Please click on the OK-button to start the ROUND 0.

<OK>

Confidence [same version in the two conditions]

Please provide your best estimate on the following question using an integer number between 0 and 100 (inclusive). You can earn additional money with your answer.

All participants in this experiment have just performed ROUND 0 under the € 0.50 piece-rate payment scheme. How many of the other participants (in percentage terms) have correctly solved LESS addition problems than you did in ROUND 0?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual percentage (rounded to the closest integer number). If your estimate deviates by at most plus or minus 1 to the actual percentage, you will earn € 4. If your estimate deviates by more than plus or minus 1 but at most plus or minus 4 from the actual percentage, you will earn € 2. If your estimate deviates by more than plus or minus 4 but at most plus or minus 7 from the actual percentage, you will earn € 1. If your estimate deviates by more than plus or minus 7 from the actual percentage, you will earn nothing.

Please fill in your best estimate here (integer number):

<OK>

Priming of group membership [GM condition]

Screen 1. Instructions Estimate dots

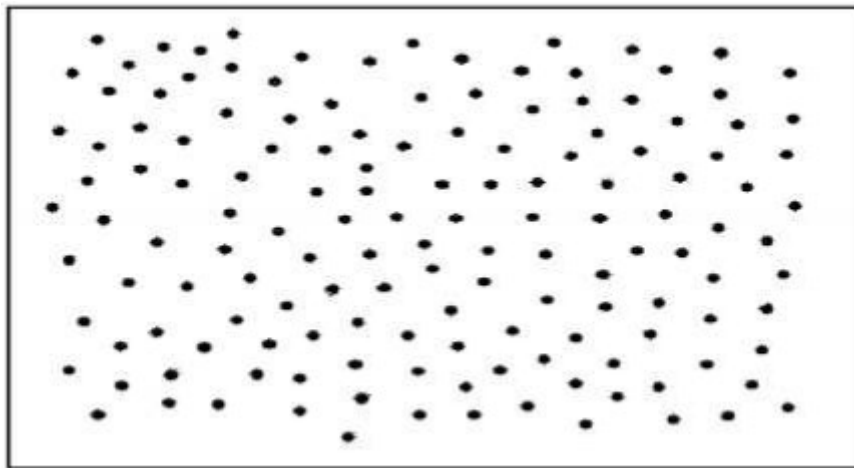
In the next screen, we display very briefly an image containing dots. Immediately after, we will ask you to guess how many dots are contained in that image.

If you have a question, please raise your hand.

Please click on the OK-button to see the image.

<OK>

Screen 2. Briefly Display of the Image (1 second)



Screen 3. Guess of Dots

How many dots are contained in the image that you saw on the previous screen?

Please fill in your best guess here (integer number):

<OK>

Screen 4. Form groups

There are 12 participants in this experimental session.

We have now formed 2 groups - Group A and Group B - of six members each. You have been assigned to GROUP A (B) with 5 other participants whose guess of dots is similar to yours.

If you wish, you can inspect the allocation rule that we use to form the two groups after the experiment.

In the next part of the experiment, the six members of Group A stay in this experimental room, whereas the six members of Group B will move to another experimental room with one of the experimenters.

Please click on the OK-button to continue.

<OK>

Assign participants to different experimental rooms [Individual condition]

There are 12 participants in this experimental session.

For the next part of the experiment, we randomly assign participants to one of two experimental rooms. Six participants stay in this experimental room. The other six participants will move to another experimental room with one of the experimenters. You have been assigned to stay in this experimental room (to move to another experimental room).

Please click on the OK-button to continue.

<OK>

Instructions for the competition rounds [GM condition]

Note: All participants are still seated in computers cubicles at Experimental Room 1. Each participant goes through four sequential screens of instructions.

Screen 1

The remaining six rounds are competition rounds. In each round, one member of Group A competes against one member of Group B. Each person in each group is randomly assigned to be the active player in one of these six rounds and to be a passive player in the five other rounds.

As the active player, you have two duties. Firstly, you choose the competition level for the round in which you compete. Secondly, you perform the addition task during 5 minutes in a one-to-one competition against the active player of the other group. We describe the details of these two roles on the next screens.

As a passive player, you have two duties. Firstly, in each of the 5 rounds, you watch in SILENCE the performance of the active player of your group (no verbal or body language are permitted!). Although you will not be able to see the active player's screen, we ask you to keep observing the active player of your group during the competition. Secondly, after the end of the last competition round, we ask you to pay attention to the feedback we provide on the choice of competition level made by each member of your group as an active player.

<Continue COMPETITION ROUNDS instructions>

Screen 2

In the round in which you are the active player, you have to solve as many addition problems as possible in a 5 minutes one-to-one competition against the active player of the other group. We randomly choose your opponent among the members of the other group. You are not informed about the identity of your opponent, neither during nor after the experiment. The same applies to your opponent.

The competition works as follows: If you correctly solve more addition problems than your opponent, you win. If you correctly solve less addition problems than your opponent, you lose. In case you and your opponent correctly solve the same number of addition problems, then there is a tie. The specific payment you receive depends not only on your performance but also on your choice of the competition level for the round in which you are the active player.

We describe on the next screen how you can choose the competition level.

<Back to the previous screen> <Continue COMPETITION ROUNDS instructions>

Screen 3

Your choice of competition level determines the exact competitive payment scheme that you face when you are the active player. Specifically, considering the following competitive payment scheme:

$$\left\{ \begin{array}{ll} \left(\frac{1}{2} + \frac{x}{100} \right) \times \text{€1} & \text{per correct answer} \quad \text{if} \quad \text{correct answers}_{\text{YOU}} > \text{correct answers}_{\text{OPPONENT}} \\ \left(\frac{1}{2} - \frac{x}{100} \right) \times \text{€1} & \text{per correct answer} \quad \text{if} \quad \text{correct answers}_{\text{YOU}} < \text{correct answers}_{\text{OPPONENT}} \end{array} \right.$$

you have to choose x belonging to $\{1, 2, 3, \dots, 49, 50\}$ to determine the exact competitive payment scheme.

Example 1: If you choose $x = 50$, you are choosing the maximum competition level possible. In this case, you perform the competition round under a winner-take-all

tournament payment scheme. That is, if you correctly solve more problems than your opponent, you win and earn $[(1/2 + 50/100) \times \text{€ } 1] = \text{€ } 1$ per correct answer. If you correctly solve less problems than your opponent, you lose and earn $[(1/2 - 50/100) \times \text{€ } 1] = \text{€ } 0$ per correct answer. In case of a tie, you earn the average of the winning and the losing payoffs, which is $(\text{€ } 1 + \text{€ } 0)/2 = \text{€ } 0.50$ per correct answer.

Example 2: If you choose $x = 1$, you are choosing the minimum competition level possible. In this case, you perform the competition round under a payment scheme in which your payment if you win is just slightly higher than your payment if you lose. That is, if you correctly solve more problems than your opponent, you win and earn $[(1/2 + 1/100) \times \text{€ } 1] = \text{€ } 0.51$ per correct answer. If you correctly solve less problems than your opponent, you lose and earn $[(1/2 - 1/100) \times \text{€ } 1] = \text{€ } 0.49$ per correct answer. In case of a tie, you earn the average of the winning and the losing payoffs, which is $(\text{€ } 0.51 + \text{€ } 0.49)/2 = \text{€ } 0.50$ per correct answer.

Example 3: If you choose $x = 25$, you are choosing an intermediate competition level. In this case, if you correctly solve more problems than your opponent, you win and earn $[(1/2 + 25/100) \times \text{€ } 1] = \text{€ } 0.75$ per correct answer. If you correctly solve less problems than your opponent, you lose and earn $[(1/2 - 25/100) \times \text{€ } 1] = \text{€ } 0.25$ per correct answer. In case of a tie, you earn the average of the winning and the losing payoffs, which is $(\text{€ } 0.75 + \text{€ } 0.25)/2 = \text{€ } 0.50$ per correct answer.

The numbers in the examples are just for illustrative purposes. They DO NOT intend to suggest how anyone should choose in this experiment.

Note: At this moment, we distribute the payoff table sheets.

<Back to the previous screen> <Continue COMPETITION ROUNDS
instructions>

Screen 4

In short: the higher you choose x , the higher is your payment if you win and the lower is your payment if you lose. The lower you choose x , the lower is your payment if you win and the higher is your payment if you lose.

IMPORTANT FINAL INFORMATION

1. When you are the active player, you are paid according to the x you choose. That is, your exact competitive payment scheme does not depend on the x chosen by your opponent.
2. You do not learn the x chosen by your opponent neither before nor after the competition round. The same applies to your opponent.

Soon, we will call all participants to stand up and go to the center of the experimental room. Before the competition rounds start, all participants see the members of each group. However, nobody in each group can see the active player

of the other group during the competition rounds because the two groups will be in separate experimental rooms. When the groups are apart, the experimenter will give you a card with a green number. The number on the card indicates the round in which you are the active player of your group.

If you understand these instructions, please click on the OK-button. Otherwise, please raise your hand. An experimenter will then come to you and answer your questions.

<Back to the previous screen> <OK>

Instructions for the competition rounds [*Individual condition*]

Note: All participants are still seated in computers cubicles at Experimental Room 1. Each participant goes through four sequential screens of instructions.

Screen 1

The remaining six rounds are competition rounds. In each round, one participant in experimental room 1 competes against one participant in experimental room 2. Each person in each room is randomly assigned to be the active player in one of these six rounds and to be a passive player in the five other rounds.

As the active player, you have two duties. Firstly, you choose the competition level for the round in which you compete. Secondly, you perform the addition task during 5 minutes in a one-to-one competition against the active player of the other room. We describe the details of these two roles on the next screens.

As a passive player, you have two duties. Firstly, in each of the 5 rounds, you watch in SILENCE the performance of the active player in your room (no verbal or body language are permitted!). Although you will not be able to see the active player's screen, we ask you to keep observing the active player in your room during the competition. Secondly, after the end of the last competition round, we ask you to pay attention to the feedback we provide on the choice of competition level made by each participant of your room as an active player.

<Continue COMPETITION ROUNDS instructions>

Screen 2

In the round in which you are the active player, you have to solve as many addition problems as possible in a 5 minutes one-to-one competition against the active player of the other room. We randomly choose your opponent among the participants in the other room. You are not informed about the identity of your opponent, neither during nor after the experiment. The same applies to your opponent.

The competition works as follows: If you correctly solve more addition problems than your opponent, you win. If you correctly solve less addition problems than

your opponent, you lose. In case you and your opponent correctly solve the same number of addition problems, then there is a tie. The specific payment you receive depends not only on your performance but also on your choice of the competition level for the round in which you are the active player.

We describe on the next screen how you can choose the competition level.

<Back to the previous screen> <Continue COMPETITION ROUNDS
instructions>

Screen 3

Your choice of competition level determines the exact competitive payment scheme that you face when you are the active player. Specifically, considering the following competitive payment scheme:

$$\begin{cases} \left(\frac{1}{2} + \frac{x}{100}\right) \times \text{€}1 & \text{per correct answer} & \text{if } \text{correct answers}_{\text{YOU}} > \text{correct answers}_{\text{OPPONENT}} \\ \left(\frac{1}{2} - \frac{x}{100}\right) \times \text{€}1 & \text{per correct answer} & \text{if } \text{correct answers}_{\text{YOU}} < \text{correct answers}_{\text{OPPONENT}} \end{cases}$$

you have to choose x belonging to $\{1, 2, 3, \dots, 49, 50\}$ to determine the exact competitive payment scheme.

Example 1: If you choose $x = 50$, you are choosing the maximum competition level possible. In this case, you perform the competition round under a winner-take-all tournament payment scheme. That is, if you correctly solve more problems than your opponent, you win and earn $[(1/2 + 50/100) \times \text{€}1] = \text{€}1$ per correct answer. If you correctly solve less problems than your opponent, you lose and earn $[(1/2 - 50/100) \times \text{€}1] = \text{€}0$ per correct answer. In case of a tie, you earn the average of the winning and the losing payoffs, which is $(\text{€}1 + \text{€}0)/2 = \text{€}0.50$ per correct answer.

Example 2: If you choose $x = 1$, you are choosing the minimum competition level possible. In this case, you perform the competition round under a payment scheme in which your payment if you win is just slightly higher than your payment if you lose. That is, if you correctly solve more problems than your opponent, you win and earn $[(1/2 + 1/100) \times \text{€}1] = \text{€}0.51$ per correct answer. If you correctly solve less problems than your opponent, you lose and earn $[(1/2 - 1/100) \times \text{€}1] = \text{€}0.49$ per correct answer. In case of a tie, you earn the average of the winning and the losing payoffs, which is $(\text{€}0.51 + \text{€}0.49)/2 = \text{€}0.50$ per correct answer.

Example 3: If you choose $x = 25$, you are choosing an intermediate competition level. In this case, if you correctly solve more problems than your opponent, you win and earn $[(1/2 + 25/100) \times \text{€}1] = \text{€}0.75$ per correct answer. If you correctly solve less problems than your opponent, you lose and earn $[(1/2 - 25/100) \times \text{€}1] =$

€ 0.25 per correct answer. In case of a tie, you earn the average of the winning and the losing payoffs, which is $(€ 0.75 + € 0.25)/2 = € 0.50$ per correct answer.

The numbers in the examples are just for illustrative purposes. They DO NOT intend to suggest how anyone should choose in this experiment.

Note: At this moment, we distribute the payoff table sheets.

<Back to the previous screen> <Continue COMPETITION ROUNDS
instructions>

Screen 4

In short: the higher you choose x , the higher is your payment if you win and the lower is your payment if you lose. The lower you choose x , the lower is your payment if you win and the higher is your payment if you lose.

IMPORTANT FINAL INFORMATION

1. When you are the active player, you are paid according to the x you choose. That is, your exact competitive payment scheme does not depend on the x chosen by your opponent.
2. You do not learn the x chosen by your opponent neither before nor after the competition round. The same applies to your opponent.

Soon, we will call all participants to stand up and go to the center of the experimental room. Before the competition rounds start, all participants see each other. However, you cannot see your opponent during the competition round because you and your opponent will be in separate experimental rooms. When the participants are apart, the experimenter will give you a card with a green number. The number on the card indicates the round in which you are the active player in your room.

If you understand these instructions, please click on the OK-button. Otherwise, please raise your hand. An experimenter will then come to you and answer your questions.

<Back to the previous screen> <OK>

4.A.3.2. Experimental instructions—Second part

Note: Half of the participants stay in experimental room 1. The other half moves to experimental room 2. In each experimental room, participants first sit in private computer cubicles to choose the competition level and answer the belief questions. For the competition rounds, the active player sits in front of a computer and the passive players sit behind the active player (but they cannot see the computer screen).

Choice of competition level [GM condition]

Note: In each computer cubicle, there is a payoff table sheet placed on the table.

ATTENTION: Your choice of competition level will determine your payment scheme in the round in which you are the active player.

Please carefully read the instructions on the next page before you take a decision.

(second page) Your choice of the competition level determines the exact competitive payment scheme that you face when you are the active player. Specifically, considering the following competitive payment scheme:

$$\left\{ \begin{array}{ll} \left(\frac{1}{2} + \frac{x}{100} \right) \times \text{€1} & \text{per correct answer} \quad \text{if} \quad \text{correct answers}_{YOU} > \text{correct answers}_{OPPONENT} \\ \left(\frac{1}{2} - \frac{x}{100} \right) \times \text{€1} & \text{per correct answer} \quad \text{if} \quad \text{correct answers}_{YOU} < \text{correct answers}_{OPPONENT} \end{array} \right.$$

you have to choose x belonging to $\{1, 2, 3, \dots, 49, 50\}$ to determine the exact competitive payment scheme that you face when you are the active player.

In short: the higher you choose x , the higher is your payment if you win and the lower is your payment if you lose. The lower you choose x , the lower is your payment if you win and the higher is your payment if you lose. Hence, a higher x implies a higher competition level.

Recall that:

1. When you are the active player, you are paid according to the x you choose. That is, your exact competitive payment scheme does not depend on the x chosen by your opponent. Moreover, you do not learn the x chosen by your opponent neither before nor after the competition round. The same applies to your opponent.
2. After the last competition round, the experimenter will reveal your choice of x to the other members of your group.

3. The other members of your group will be watching you while you compete. Still, they will neither see the addition problems you are tackling during the competition round, nor do they learn at any moment of the experiment whether you win or lose the competition.

In addition, we would like to inform you that EACH of the other members of your group earns an amount equal to $1/3$ of your earnings in the round in which you compete. For example, if you earn € 12 in the competition, we not only pay you € 12 but also pay $1/3 \times € 12 = € 4$ to EACH of the other members of your group.

Please make your choice of competition level (integer number between 1 and 50, inclusive)

Choice of competition level [Individual condition]

Note: In each computer cubicle, there is a payoff table sheet placed on the table.

ATTENTION: Your choice of competition level will determine your payment scheme in the round in which you are the active player.

Please carefully read the instructions on the next page before you take a decision.

(second page) Your choice of the competition level determines the exact competitive payment scheme that you face when you are the active player. Specifically, considering the following competitive payment scheme:

$$\left\{ \begin{array}{ll} \left(\frac{1}{2} + \frac{x}{100} \right) \times €1 & \text{per correct answer} \quad \text{if} \quad \text{correct answers}_{YOU} > \text{correct answers}_{OPPONENT} \\ \left(\frac{1}{2} - \frac{x}{100} \right) \times €1 & \text{per correct answer} \quad \text{if} \quad \text{correct answers}_{YOU} < \text{correct answers}_{OPPONENT} \end{array} \right.$$

you have to choose x belonging to $\{1, 2, 3, \dots, 49, 50\}$ to determine the exact competitive payment scheme that you face when you are the active player.

In short: the higher you choose x , the higher is your payment if you win and the lower is your payment if you lose. The lower you choose x , the lower is your payment if you win and the higher is your payment if you lose. Hence, a higher x implies a higher competition level.

Recall that:

1. When you are the active player, you are paid according to the x you choose. That is, your exact competitive payment scheme does not depend on the x chosen by your opponent. Moreover, you do not learn the x chosen by your opponent neither before nor after the competition round. The same applies to your opponent.
2. After the last competition round, the experimenter will reveal your choice of x to the other participants in your room.
3. The other participants in your room will be watching you while you compete. Still, they will neither see the addition problems you are tackling during the competition round, nor do they learn at any moment of the experiment whether you win or lose the competition.

Please make your choice of competition level (integer number between 1 and 50, inclusive)

First-order beliefs about the choice of competition level [*GM condition*]

Please provide your best estimate on the following question using an integer number between 1 and 50 (inclusive). You can earn additional money with your answer.

Consider the other members of your group. What is your estimate of their choice of competition level, on average?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual percentage (rounded to the closest integer number). If your estimate deviates by at most plus or minus 1 to the actual percentage, you will earn € 2. If your estimate deviates by more than plus or minus 1 but at most plus or minus 4 from the actual percentage, you will earn € 1. If your estimate deviates by more than plus or minus 4 but at most plus or minus 7 from the actual percentage, you will earn € 0.5. If your estimate deviates by more than plus or minus 7 from the actual percentage, you will earn nothing.

Please fill in your best estimate here (integer number):

First-order beliefs about the choice of competition level [*Individual condition*]

Please provide your best estimate on the following question using an integer number between 1 and 50 (inclusive). You can earn additional money with your answer.

Consider the other participants of your room. What is your estimate of their choice of competition level, on average?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual percentage (rounded to the closest integer number). If your estimate deviates by at most plus or minus 1 to the actual percentage, you will earn € 2. If your estimate deviates by more than plus or minus 1 but at most plus or minus 4 from the actual percentage, you will earn € 1. If your estimate deviates by more than plus or minus 4 but at most plus or minus 7 from the actual percentage, you will earn € 0.5. If your estimate deviates by more than plus or minus 7 from the actual percentage, you will earn nothing.

Please fill in your best estimate here (integer number):

Second-order beliefs about the choice of competition level [GM condition]

Please provide your best estimate on the following question using an integer number between 1 and 50 (inclusive). You can earn additional money with your answer.

Consider the other members of your group. What is your estimate of their average estimate of your choice of competition level?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual percentage (rounded to the closest integer number). If your estimate deviates by at most plus or minus 1 to the actual percentage, you will earn € 2. If your estimate deviates by more than plus or minus 1 but at most plus or minus 4 from the actual percentage, you will earn € 1. If your estimate deviates by more than plus or minus 4 but at most plus or minus 7 from the actual percentage, you will earn € 0.5. If your estimate deviates by more than plus or minus 7 from the actual percentage, you will earn nothing.

Please fill in your best estimate here (integer number):

Second-order beliefs about the choice of competition level [Individual condition]

Please provide your best estimate on the following question using an integer number between 1 and 50 (inclusive). You can earn additional money with your answer.

Consider the other participants of your room. What is your estimate of their average estimate of your choice of competition level?

Your payment will be higher, the better your estimate. Specifically, your payment will be higher the closer your estimate is to the actual percentage (rounded to the closest integer number). If your estimate deviates by at most plus or minus 1 to the actual percentage, you will earn € 2. If your estimate deviates by more than plus or minus 1 but at most plus or minus 4 from the actual percentage, you will earn € 1. If your estimate deviates by more than plus or minus 4 but at most plus or minus 7 from the actual percentage, you will earn € 0.5. If your estimate deviates by more than plus or minus 7 from the actual percentage, you will earn nothing.

Please fill in your best estimate here (integer number):

Feedback on other's choice of competition level [*same version in the two conditions*]

Note: In each experimental room, the feedback is provided right after the last competition round is over.

ACTIVE PLAYERS choice of competition level

Round 1 #_____

Round 2 #_____

Round 3 #_____

Round 4 #_____

Round 5 #_____

Round 6 #_____

4.A.3.2. Experimental instructions—Third part

Note: All participants again sit in private computer cubicles at experimental room 1.

Instructions for the elicitation of risk preference—end [GM condition]

In the following screen, you will have to again make a choice between a lottery and a sure payment in 15 decision situations. The lottery is the same across the 15 decision situations, but the sure payment is always different.

Once you have made a choice for each one of the 15 decision situations the computer randomly selects one decision situation for payment. Your earnings will depend on your choice for that decision situation. If you have chosen the lottery for that decision situation, you will earn with a 50 percent chance 400 points and with a 50 percent chance 0 points. If you have chosen the sure payment for that decision situation, you will earn the sure payment corresponding to that decision situation.

In addition, we would like to inform you that EACH of the other members of your group earns an amount equal to $1/3$ of your earnings. For example, if you earn 300 points in the decision situation the computer randomly selects for payment, we not only pay you 300 points but also pay $1/3 \times 300 = 100$ points to EACH of the other members of your group.

The points you earn in this part of the experiment will be exchanged into Euros at the end of the experiment with an exchange rate of:

100 points = € 0.50

If you have a question, please raise your hand.

Please click CONTINUE to make your choices.

<CONTINUE>

Instructions for the elicitation of risk preference—end [Individual condition]

Exactly as at the start.

Elicitation of risk preference—end [*GM condition*]

Please decide for each decision situation whether you want to choose the lottery or the sure payment!
Each of the other members of your group earns an amount equal to 1/3 of your earnings.

Decision Situation	Lottery	Sure payment	Your choice
1	50% chance of 400 points and 50% chance of 0 points	25 points	lottery <input type="radio"/> sure payment <input type="radio"/>
2	50% chance of 400 points and 50% chance of 0 points	50 points	lottery <input type="radio"/> sure payment <input type="radio"/>
3	50% chance of 400 points and 50% chance of 0 points	75 points	lottery <input type="radio"/> sure payment <input type="radio"/>
4	50% chance of 400 points and 50% chance of 0 points	100 points	lottery <input type="radio"/> sure payment <input type="radio"/>
5	50% chance of 400 points and 50% chance of 0 points	125 points	lottery <input type="radio"/> sure payment <input type="radio"/>
6	50% chance of 400 points and 50% chance of 0 points	150 points	lottery <input type="radio"/> sure payment <input type="radio"/>
7	50% chance of 400 points and 50% chance of 0 points	175 points	lottery <input type="radio"/> sure payment <input type="radio"/>
8	50% chance of 400 points and 50% chance of 0 points	200 points	lottery <input type="radio"/> sure payment <input type="radio"/>
9	50% chance of 400 points and 50% chance of 0 points	225 points	lottery <input type="radio"/> sure payment <input type="radio"/>
10	50% chance of 400 points and 50% chance of 0 points	250 points	lottery <input type="radio"/> sure payment <input type="radio"/>
11	50% chance of 400 points and 50% chance of 0 points	275 points	lottery <input type="radio"/> sure payment <input type="radio"/>
12	50% chance of 400 points and 50% chance of 0 points	300 points	lottery <input type="radio"/> sure payment <input type="radio"/>
13	50% chance of 400 points and 50% chance of 0 points	325 points	lottery <input type="radio"/> sure payment <input type="radio"/>
14	50% chance of 400 points and 50% chance of 0 points	350 points	lottery <input type="radio"/> sure payment <input type="radio"/>
15	50% chance of 400 points and 50% chance of 0 points	375 points	lottery <input type="radio"/> sure payment <input type="radio"/>

Please click OK if you have made your choice for each decision situation!

OK

Elicitation of risk preference—end [*Individual condition*]

Exactly as at the start.

Chapter 5

Synopsis of the Thesis

This thesis is the compilation of three studies on competitive behavior. More specifically, I investigate the role of sex, stereotypes, and group membership in competitive settings. This chapter presents summaries of each study.

5.1. Women dislike competing against men

The literature on gender competition depicts women as less inclined to compete than men. In **Chapter 2**, I conduct a laboratory experiment to examine the relevance of beliefs about the sex of potential competitors on men's and women's decision to enter competition. Specifically, I test whether women have a weaker preference to compete, or whether any realized preference varies by the sex of the competitor.

The results support the latter hypothesis. That is, when given the possibility of choosing a competitor's sex, or when being in the lab surrounded only by female participants, the percentage of women entering competition is high and similar to the figures commonly reported for men. Moreover, only women are sensitive to the different cues we provide about the sex of potential competitors, and their competitiveness is largely driven by their beliefs about other women's competitive attitude.

These findings have distinctive policy implications for the labor markets in which women are underrepresented. Notably, I argue that on-going interventions that highlight women's underrepresentation in job advertisements, which are intended to encourage women to apply, could be triggering the opposite effect.

Instead, I advocate a nudge in the form of persuasive references to recent female applicants and/or existing female staff.

5.2. Sex and performance under competition: Is there a stereotype threat shadow? An intriguing behavior of men¹

In **Chapter 3**, we conduct laboratory and online experiments to study the competitive performance of men and women across conditions in which we implicitly or explicitly induce a stereotype threat. To this end, we use a mathematical task that implicitly creates a stereotype threat against women. We also study conditions in which we explicitly reinforce or contradict the implicit stereotype threat by providing appropriate information.

Our findings contrast previous results suggesting that men perform better than women under competition. Firstly, men and women react positively and equally strong to competitive incentives when we induce stereotype threats against women. Secondly, only women respond positively to competitive incentives when we explicitly induce a stereotype threat against men.

In this chapter, we also study the effort a participant should strategically exert in our one-to-one competitive setting. We find that when beliefs in a considerable gap in ability between a pair of competitors are common knowledge, the equilibrium effort of the more able competitor is higher than the effort level he/she believes is sufficient to win the competition.

Collectively, our results have practical implications for policy design directed at stereotype threats in the workplace. Specifically, in contexts where men and women are exogenously set to compete against each other, the appropriate intervention to prevent the adverse effect of stereotype threat in performance is to avoid any information making reference to the stereotype.

5.3. Group membership and individual competitive behavior

Previous studies have shown that, in non-competitive settings, individuals who feel part of a group behave differently compared to individuals who perceive themselves as isolated. In **Chapter 4**, I conduct a laboratory experiment to test if (and how) group membership affects individual willingness to compete. For this purpose, I contrast a context in which I induce participants to feel part of a group to a context in which participants are on their own. Also, to better measure willingness to compete, I design a new indicator that allows the measurement of several different levels of competitiveness.

I find that participants induced with group membership choose significantly lower levels of competition compared to control participants. Interestingly, I find that group members' competitiveness is driven by their belief about other group

¹ This chapter is based on Gerdle, Riedl, and Strobel (2016)

members' expectations of the group's degree of competitiveness. I argue that group membership restrains competitive behavior in my setting (partially) because of guilt aversion.

Moreover, even though members of a group choose lower competitive levels, I find that they perform better under competition relative to control participants. I attribute the latter result to the presence of an audience during the competition, which is composed of the other members of the group in the group membership context.

The findings in this chapter suggest two thought-provoking policies to improve decision-making in competitive labor markets. Firstly, to deter reckless decisions, we should index the bonus of workers to the performance of their peers. Secondly, to improve workers' productivity, we should strengthen their group spirit to compete against external organizations rather than promoting internal competition.

Valorization Addendum

Knowledge valorization refers to the “process of creating value from knowledge, by making knowledge suitable and/or available for social (and/or economic) use and by making knowledge suitable for translation into competitive products, services, processes and new commercial activities” (adapted definition based on the National Valorisation Committee 2011:8).

*In: Regulations governing the attainment of doctoral degrees
Maastricht University (2013)*

Policy recommendations

This dissertation addresses valorization itself. Throughout my thesis, I draw policy implications for the real-world based on my research. In this addendum, to avoid duplication, I only summarize the key policy recommendations contained in this dissertation:

➤ I prescribe that highlighting female role models can increase women’s willingness to enter competitive male-typed domains. Namely, I propose that information in job advertisements that highlight women’s underrepresentation—which is, for instance, a common practice in the economics academic job market—should be replaced by references to recent female applicants and/or existing female workers. (Chapter 2)

➤ Regarding policy design to cope with stereotype threat at competitive workplaces, my results indicate that—if men and women are already competing against each other—the appropriate intervention to prevent the adverse effect of stereotype threat in performance is to avoid any information making reference to the stereotype. (Chapter 3)

➤ To deter reckless decisions in competitive labor markets, we should index the bonus of workers to the performance of their peers. This incentive scheme could be applied, for instance, to tackle the infamous problem of irresponsible investment in the hedge fund world. (Chapter 4)

Contributions to public debate and promotion of Maastricht University

The research contained in this dissertation has been disseminated in prestigious places such as, for instance, the Yale University Whitebox Conference, Econometric Society European Winter Meeting, Stanford Institute for Theoretical Economics Summer Workshop, Columbia University Political Economy Seminar, and the NHH Choice Lab Seminar. In total, the research output contained in this dissertation originated more than twenty oral presentations under the label of Maastricht University.

Also, the research underlying Chapter 2 has received the best paper award at the Gender Gaps Conference under the label of Maastricht University. This achievement was recognized with a research excellence distinction from the Dean of the Maastricht University School of Business and Economics at the 2017 End-Of-Year SBE event.

Software

The z-Tree code written to computerize the experiments underlying the research presented in Chapter 2, Chapter 3, and Chapter 4 is available upon request at <https://sites.google.com/site/diogogeraldes>.

Academic Biography

Diogo José da Cunha Batista Geraldês was born in São Paulo (Brazil) to Portuguese parents, and brought to Portugal with seven months. Diogo graduated from high school at Escola Secundária José Falcão in Coimbra. At the age of 18, Diogo moved to Lisboa to study economics at Nova SBE, where he graduated in the top 3% of his cohort (class 2003). During his B.Sc. in economics, Diogo visited the London School of Economics and Political Science. Also, during his last year as an undergrad student, Diogo received an invitation to attend the Nova SBE Ph.D. in Economics program, which start Diogo postponed for taking a gap year in Brussels.

In September 2004, Diogo started the Ph.D. in economics program at Nova SBE (fully funded). While in the second year of the program—where Diogo obtained the best preliminary Ph.D. entry examination in macroeconomics—Diogo was highly encouraged by his Nova SBE professors to apply for Ph.D. in economics programs outside Portugal.

In September 2006, Diogo started the Ph.D. in economics program at Pompeu Fabra University (fully funded) aiming to write a dissertation in political economy. After fulfilling the first-year coursework requirements and successfully starting a *Tesina* (which is a draft of the first paper in a Ph.D. dissertation) in political economy, Diogo was among the few students selected to stay for the second-year.

During his second-year at Pompeu Fabra, a radical change in Diogo's research orientation happened. After taking the *Behavioral Decision Making* course of Prof. Dr. Robin Hogarth and the *Experimental Economics* course of Prof. Dr. Rosemarie Nagel, Diogo decided to rather write his *Tesina* in behavioral and experimental economics. This decision, which was deemed highly risky and counter prudent by the program directors, led in the space of only a few months to the *Tesina* entitled *How do we operate under Ambiguity: An Experimental Approach*, which was one of the few approved *Tesinas* in that academic year, with a pre-classification of 9 (on a 0-10 scale). Also in his second-year at Pompeu Fabra, Diogo wrote a term paper about neuroeconomics—which also obtained a classification of 9—for the *Behavioral Decision Making* course. This latter experience triggered Diogo's interest in a Ph.D. position in neuroeconomics at Maastricht University.

In July 2008, Diogo accepted an offer to proceed his Ph.D. at Maastricht University with a major focus on neuroeconomics. From October 2008, Diogo conducted research directed to neuroeconomics for 3 years, intermediated by an unfortunate sick leave due to a retinal detachment. Alas, a multiplicity of factors turned the planned neuroeconomics project unsustainable. As result, in the Winter

term of 2011/2012, Diogo decided—with the acceptance of his supervisors—to drop the neuroeconomics project and write his Ph.D. dissertation exclusively in behavioral and experimental economics.

In the research presented in this dissertation, Diogo investigates competitive settings, particularly gender competition and the role of group membership, and the policy implications for the labor markets. Part of this dissertation was conducted at Columbia University, where Diogo held a visiting scholar position. Diogo's research has been presented at various places such as, for instance, Yale University, Stanford University, Columbia University, and the NHH Choice Lab Seminar. Also, the paper underlying Chapter 2—"Women dislike competing against men"—received the best paper award at the Gender Gaps Conference.

Since September 2017, Diogo holds the position of assistant professor in the Chair of Microeconomics at the Utrecht School of Economics.